

|  |  |
|--|--|
| Scoring table:                                     | IDENTITY_NUC                                 |
| Title:   | US-09-673-716-1                              |
| Perfect score:                                     | 231  |
| Sequence:  | 1 gtgggtgcggatctaa...ttcacccataggcttgctt 231 |
| Total number of hits satisfying chosen parameters: | 45562784                                     |
| Minimum DB seq length:                             | 0  |
| Maximum DB seq length:                             | 2000000000                                   |
| Post-processing:                                   | Minimum Match 0%                             |
|  | Maximum Match 100%                           |
|  | Listing first 45 summaries                   |
| Database :   | EST:*  |
| 1:   | em_estba:*                                   |
| 2:   | em_estthm:*                                  |
| 3:   | em_estin:*                                   |
| 4:   | em_estinu:*                                  |
| 5:   | em_estov:*                                   |
| 6:   | em_estpl:*                                   |
| 7:   | em_estro:*                                   |
| 8:   | em_htc:*                                     |
| 9:   | gb_est1:*                                    |
| 10:  | gb_est2:*                                    |
| 11:  | gb_htc:*                                     |
| 12:  | qb_est3:*                                    |
| 13:  | qb_est4:*                                    |
| 14:  | qb_est5:*                                    |
| 15:  | em_estin:*                                   |
| 16:  | em_estom:*                                   |
| 17:  | em_gss_hum:*                                 |
| 18:  | em_gss_inv:*                                 |
| 19:  | em_gss_pln:*                                 |
| 20:  | em_gss_vrt:*                                 |
| 21:  | em_gss_fam:*                                 |
| 22:  | em_gss_mam:*                                 |
| 23:  | em_gss_mus:*                                 |
| 24:  | em_gss_pro:*                                 |
| 25:  | em_gss_rid:*                                 |
| 26:  | em_gss_phg:*                                 |
| 27:  | em_gss_vrl:*                                 |
| 28:  | gb_gss1:*                                    |
| 29:  | gb_gss2:*                                    |
| 5:   | 231 100.0 559 28 BH105759 RPCI-24-2          |
| c 6:   | 231 100.0 577 28 AZ20721 1M0198L07           |
| c 7:   | 231 100.0 587 28 AZ832364 2M0112B13          |
| c 8:   | 231 100.0 627 28 AZ95883 1M0332E03           |
| c 9:   | 231 100.0 647 28 AZ929390 2M0279D16          |
| c 10:  | 231 100.0 706 14 CB056516 NISC_JJ14          |
| c 11:  | 231 100.0 721 28 AZ94818 2M0280J4            |
| c 12:  | 229 4 99.3 472 28 AZ310386 1M002503          |
| c 13:  | 229 4 99.3 505 28 AZ054649 RPCI-23-4         |
| c 14:  | 229 4 99.3 518 9 AW58986 L0302F10-           |
| c 15:  | 229 4 99.3 520 12 BM219777 C0929H06-         |
| c 16:  | 229 4 99.3 534 28 AZ311190 1M0073N4          |
| c 17:  | 229 4 99.3 662 28 AZ326140 1M0048D11         |
| c 18:  | 229 4 99.3 535 28 AZ965228 2M0235A17         |
| c 19:  | 229 4 99.3 539 28 AW554086 L0235F05-         |
| c 20:  | 229 4 99.3 557 9 AW554086                    |
| c 21:  | 229 4 99.3 568 28 AZ987016 2M0269M43         |
| c 22:  | 229 4 99.3 584 28 AZ312413 1M0217N95         |
| c 23:  | 229 4 99.3 588 28 AZ22504 RPCI-23-8          |
| c 24:  | 229 4 99.3 678 14 BY31717 BY731717           |
| c 25:  | 229 4 99.3 686 28 BH092157 RPCI-24-2         |
| c 26:  | 229 4 99.3 707 28 BH117452 RPCI-24-2         |
| c 27:  | 229 4 99.3 834 28 BH062187 RPCI-24-3         |
| c 28:  | 229 4 99.3 4484 11 AK033082 Mus muscu        |
| c 29:  | 228 4 98.9 606 12 B1134544 UI-M-BH3-         |
| c 30:  | 227 8 98.6 309 28 BH068706 RPCI-24-3         |
| c 31:  | 227 8 98.6 457 28 AZ071850 RPCI-23-4         |
| c 32:  | 227 8 98.6 459 9 AW52509 L0213E08-           |
| c 33:  | 227 8 98.6 531 28 A2636367 1M0495J10         |
| c 34:  | 227 8 98.6 557 28 BH078165 RPCI-24-3         |
| c 35:  | 227 8 98.6 559 13 BQ53681 H4023F02-          |
| c 36:  | 227 8 98.6 596 28 A2830822 2M0110H4          |
| c 37:  | 227 8 98.6 616 28 BH099567 RPCI-24-3         |
| c 38:  | 227 8 98.6 678 28 A2897416 RPCI-24-2         |
| c 39:  | 227 8 98.6 686 28 A2259805 RPCI-23-1         |
| c 40:  | 227 8 98.6 706 28 A2797045 2M0053J06         |
| c 41:  | 227 8 98.6 793 28 A2719135 RPCI-24-1         |
| c 42:  | 227 8 98.6 838 28 BH027115 RPCI-24-2         |
| c 43:  | 227 4 98.4 553 28 AZ792711 2M0045C19         |
| c 44:  | 227 4 98.4 642 28 AZ090071 RPCI-23-4         |
| c 45:  | 227 4 98.4 675 28 AZ957637 2M0224N06         |

## ALIGNMENTS

| RESULT           | 1   | AZ231159    | LOCUS   | RPCI-23-51R23 | 477 bp DNA                 | DEFINITION               | Mus musculus genomic clone | GSS 14-JUN-2000 |
|------------------|---|-------------|---|---------------|----------------------------|--------------------------|----------------------------|-----------------|
|                  |   |             |   |               |                            | genomic survey sequence. |                            |                 |
| ACCESSION        | AZ231159  |             | VERSION   | AZ231159      |                            | KEYWORDS                 | GSS                        |                 |
| SOURCE           |   |             | ORGANISM  |               | Mus musculus (house mouse) |                          |                            |                 |
| ORGANISM         |   |             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus. |               |                            |                          |                            |                 |
| REFERENCE        |   |             | 1 (bases 1 to 477)  |               |                            |                          |                            |                 |
| AUTHORS          | Zhao S., Niemann W., Feldblum T., Malek J., Shatsman S., Akinret, B., Levins M., McGann S., Tsegaye G., Geer K., Kroli M., de Jong P. |             |   |               |                            |                          |                            |                 |
| TITLE            |   |             | Mouse BAC End Sequences from Library RPCI-23  |               |                            |                          |                            |                 |
| JOURNAL          |   |             | Unpublished Other GSS: RPCI-23-51K23.TV   |               |                            |                          |                            |                 |
| COMMENT          |   |             | Contact: Shaying Zhao   |               |                            |                          |                            |                 |
|                  |   |             | Department of Eukaryotic Genomics   |               |                            |                          |                            |                 |
|                  |   |             | The Institute for Genomic Research  |               |                            |                          |                            |                 |
|                  |   |             | 9712 Medical Center Dr., Rockville, MD 20850, USA   |               |                            |                          |                            |                 |
|                  |   |             | Tel: 301 838 0200   |               |                            |                          |                            |                 |
|                  |   |             | Fax: 301 838 0208   |               |                            |                          |                            |                 |
|                  |   |             | Email: szhao@tigr.org   |               |                            |                          |                            |                 |
|                  |   |             | Clones are derived from the mouse BAC library RPCI-23. For BAC  |               |                            |                          |                            |                 |
| <b>SUMMARIES</b> |   |             |   |               |                            |                          |                            |                 |
| Result No.       | Score   | Query Match | Length  | DB ID         | Description                |                          |                            |                 |
| 1                | 231 100.0   | 477         | 28  | AZ231159      |                            |                          |                            |                 |
| 2                | 231 100.0   | 501         | 28  | Az440299      | AZ440299 1M0231K17         |                          |                            |                 |
| 3                | 231 100.0   | 527         | 10  | BG087407      | EG087407 H139B12-          |                          |                            |                 |
| c 4              | 231 100.0   | 558         | 28  | BH036520      | BH036520 RPCI-24-2         |                          |                            |                 |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://inforegen.com>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)

Plate: 51 row: K column: 23

Seq primer: SP6

Class: BAC ends.

#### FEATURES

##### source

1. 477  
*/lab\_host="DH10B"  
/clone\_lib="RPCI-23"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1:  
EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and ECOLI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies).*  
  
BASE COUNT 94 a  
ORIGIN 110 c 143 g 130 t

Query Match 100.0%; Score 231; DB 28; Length 477;  
Best Local Similarity 100.0%; Pred. No. 1..5e-60;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGGACTGACAGAGGATAGCTCTGTGGCATCCTGTGGAAAG 60  
Db 232 GTGGGTGCGAGGCTAAGGACTGACAGAGGATAGCTCTGTGGCATCCTGTGGAAAG 291  
QY 61 CACGCTGATTGCAAGTGAAGGTCAGTGCTTAGTGTCCCTCCCAGAAAGACACG 120  
Db 292 CACGCTGATTGCAAGTGAAGGTCAGTGCTTAGTGTCCCTCCCAGAAAGACACG 351

QY 121 GGAGCTGCCAAGAACCTCTGGGTATAGCCATAAGGGATGGTTTGTAGGGCCCT 180  
Db 352 GGAGCTGCCAAGAACCTCTGGGTATAGCCATAAGGGATGGTTTGTAGGGCCCT 411  
QY 181 ATGCCTGCACACTGGGATCAGACTCACCTCACCCATGAGCTCTGCTT 231  
Db 412 ATGCCTGCACACTGGGATCAGACTCACCTCACCCATGAGCTCTGCTT 462

RESULT 2  
AZ440299 LOCUS AZ440299 501 bp DNA linear GSS 03-OCT-2000  
DEFINITION IM0231K17F Mouse 10kb plasmid UGCLM library Mus musculus genomic  
clone UGCLM0231K17 F, genomic survey sequence.  
ACCESSION AZ440299  
VERSION AZ440299\_1 GI:10564312  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus.

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud.M., Meenah,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-51K23"  
/sex="Female"  
/lab\_host="DH10B"  
/clone\_lib="RPCI-23"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1:  
EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and ECOLI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

##### FEATURES

##### source

1. 501  
*/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb Plasmid UGCLM library"  
/note="Vector: pND42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnases/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pND42 (91473211 [gb] AF129072.1), a copy-number  
inducible derivative of Plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor-modified mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."*

RESULTS 3  
LOCUS BG087407  
DEFINITION H313912-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
VERSION H313912-5 , mRNA sequence.  
ACCESSION BG087407  
KEYWORDS EST  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Rodentia; Muridae; Murinae; Mus.  
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islami,H., Longacre,S., Mahmoud.M., Meenah,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
PAGE 2

|                            |   |    |   |
|----------------------------|---|----|---|
| REFERENCE                  | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.<br>1 (bases 1 to 527)  | Qy | 181 ATGGCTTGACACTGGGATCAGACCTTACCCATGAGGCTTGCTT 231             |
| AUTHORS                    | Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac<br>,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.<br>III., Becker,K.G. and Ko,M.S.H.   | Db | 416 ATGGCTTGACACTGGGATCAGACCTTACCCATGAGGCTTGCTT 466             |
| TITLE                      | Genome-wide expression profiling of mid-gestation placenta and<br>embryo using a 15,000 mouse developmental cDNA microarray   |    |   |
| JOURNAL                    | Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)   |    |   |
| MEDLINE                    | 10922068  |    |   |
| PUBLISHED                  | Other ESTs: H3139B12-3  |    |   |
| COMMENT                    | Contact: George J. Kargul<br>Laboratory of Genetics<br>National Institute on Aging/National Institutes of Health<br>333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA<br>Email: cdna@igsun.grc.nih.gov<br>This clone set has been freely distributed to the community. Please<br>visit <a href="http://igsun.grc.nia.nih.gov/cDNA/15k.html">http://igsun.grc.nia.nih.gov/cDNA/15k.html</a> for details.<br>Plate: H3139 Row: B Column: 12  |    |   |
| SEQ PRIMER                 | -21M13 Reverse  |    |   |
| HIGH QUALITY SEQUENCE STOP | 527   |    |   |
| POLYA-NO.                  | Location/Qualifiers   |    |   |
| FEATURES                   | 1. -527<br>/organism="Mus musculus"<br>/mol_type="mRNA"<br>/strain="C57BL/6J"<br>/db_xref="NIAEST:H3139B12-5"<br>/db_xref="taxon:10090"<br>/clone="H3139B12"<br>/sex="Clones arrayed from a variety of cDNA libraries"<br>/dev_stage="Clones arrayed from a variety of cDNA<br>libraries"<br>/lab_host="DH10B"<br>/clone.lib="NIA Mouse 15K cDNA Clone Set"<br>/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This<br>clone is among a rearrayed set of 15,247 clones from 11<br>embryo cDNA libraries (including preimplantation stage<br>embryos from unfertilized egg to blastocyst, embryonic<br>part of E7.5 embryos, extraembryonic part of E7.5 embryos<br>, and E12.5 female mesonephros/gonad, and one newborn<br>ovary cDNA library. Average insert size 1.5 kb. All<br>source libraries are cloned unidirectionally with Oligo(dT<br>)-Not primers. References include: (1) Genome-wide<br>expression profiling of mid-gestation placenta and embryo<br>using a 15,000 mouse developmental cDNA microarray, 2000,<br>Proc. Natl. Acad. Sci. U.S.A., 97: 9127-9132; (2)<br>Large-scale cDNA analysis reveals phased gene expression<br>patterns during preimplantation mouse development, 2000,<br>Development, 127: 1731-1749; (3) Genome-wide mapping of<br>unselected transcripts from extraembryonic tissue of<br>7.5-day mouse embryos reveals enrichment in the t-complex<br>and under-representation on the X chromosome, 1998, Hum<br>Mol Genet 7: 1967-1978." |    |   |
| source                     |   |    |   |
| BASE COUNT                 | 104 a 121 c 158 g 144 t   |    |   |
| ORIGIN                     |   |    |   |
| Query Match                | 100.0%; Score 231; DB 10; Length 527;   |    |   |
| Best Local Similarity      | 100.0%; Pred. No. 1. 6e-60;   |    |   |
| Matches                    | 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |    |   |
| Qy                         | 1 GTGGGGTGGGAGGCTAAGGACTGGACAGAGGATAGCTTGTGGCATCCCTGGAGG 60   | Qy | 1 GTGGGGTGGGAGGCTAAGGACTGGACAGAGGATAGCTTGTGGCATCCCTGGAGG 60     |
| Db                         | 236 GTGGGGTGGGAGGACTGGACAGAGGATAGCTTGTGGCATCCCTGGAGG 295  | Db | 376 GTGGGGTGGGAGGCTAAGGACTGGACAGAGGATAGCTTGTGGCATCCCTGGAGG 317  |
| Qy                         | 61 CACGCTCTGATTGCAATGAGGTTCAAGTGTCTAGTCCTCCCTCCAGAAACGACAG 120  | Qy | 61 CACGCTCTGATTGCAATGAGGTTCAAGTGTCTAGTCCTCCCTCCAGAAACGACAG 120  |
| Db                         | 296 CACGCTCTGATTGCAATGAGGTTCAAGTGTCTAGTCCTCCCTCCAGAAACGACAG 355   | Db | 316 CACGCTCTGATTGCAATGAGGTTCAAGTGTCTAGTCCTCCCTCCAGAAACGACAG 257 |
| Qy                         | 121 GGACCTGGCCAAGACCTCTCTGGGTGAAGGCTTAAGGATGGTTGTAGGCCCT 180  | Qy | 121 GGACCTGGCCAAGACCTCTCTGGGTGAAGGCTTAAGGATGGTTGTAGGCCCT 180    |
| Db                         | 256 GGACCTGGCCAAGACCTCTCTGGGTGAAGGCTTAAGGATGGTTGTAGGCCCT 197  | Db |   |

|                       |  |  |   |         |        |                 |
|-----------------------|--|--|---|---------|--------|-----------------|
| Qy                    | 181  | ATGCTTGACACTGGGATCAGACCTTACCCCTAACGCTTACCCATGAGCTGCT   | 231   |         |        |                 |
| Db                    | 196  | ATGCTTGACACTGGGATCAGACCTTACCCCTAACGCTTACCCATGAGCTGCT   | 146   |         |        |                 |
| RESULT 5              | BH105759   | BH105759   | 559 bp  | DNA     | linear | GSS 19-JUL-2001 |
| DEFINITION            | RPCI-24-237C16 TU RPCI-24 Mus musculus genomic clone RPCI-24-237C16 genomic survey sequence.   |  |   |         |        |                 |
| ACCESSION             | BH105759   | BH105759   |   |         |        |                 |
| VERSION               | 1  | GT:14935045  |   |         |        |                 |
| KEYWORDS              | GSS.   | Mus musculus (house mouse)   |   |         |        |                 |
| ORGANISM              | Mus musculus   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |   |         |        |                 |
| REFERENCE             | 1  | (bases 1 to 559)   |   |         |        |                 |
| AUTHORS               | Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsehayye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregheorgis,E., Russell,D., de Jong,P., and Fraser,C.M.   | Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Dural,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Relly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D. Weiss,R.   |   |         |        |                 |
| TITLE                 | Mouse BAC End Sequences from Library RPCI-24   | Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts   |   |         |        |                 |
| JOURNAL               | Unpublished  | Unpublished  |   |         |        |                 |
| COMMENT               | Contact: Shaying Zhao<br>Department of Eukaryotic Genomics<br>The Institute for Genomic Research<br>9712 Medical Center Dr., Rockville, MD 20850, USA<br>Tel.: 301 838 0200<br>Fax: 301 838 0208<br>Email: szhao@igb.org | Contact: Robert B. Weiss<br>University of Utah Genome Center<br>University of Utah<br>Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT<br>Tel: 801 585 5606<br>Fax: 801 585 7177<br>Email: ddunn@genetics.utah.edu<br>Insert Length: 10000 Std Error: 0.00<br>Plate: 0198 row: L column: 07<br>Seq primer: CACACGGAAACAGCTATGACC<br>Class: Plasmid ends<br>High quality sequence stop: 577. |   |         |        |                 |
| FEATURES              | source   | Location/Qualifiers  |   |         |        |                 |
|                       |  | 1. .559  | /organism="Mus musculus"<br>/mol_type="genomic DNA"<br>/strain="C57BL/6J"<br>/db_xref="taxon:10090"<br>/clone "RPCI-24-237C16"<br>/sex="Male"<br>/cell_type="Spleen/Brain"<br>/clone_id="RPCI-24"<br>/note="vector: PTBAC1; Site:1: BamHI; Site:2: BamHI;<br>RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the PTBAC1 Cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA." |         |        |                 |
|                       |  | 119 a  | 132 c   | 159 g   | 149 t  |                 |
| BASE COUNT            |  |  |   |         |        |                 |
| ORIGIN                |  |  |   |         |        |                 |
| Query Match           | 100.0%   | Score 231; DB 28;  | Length 559;   |         |        |                 |
| Best Local Similarity | 100.0%   | Pred. No. 1..6e-60;  |   |         |        |                 |
| Matches 231:          | Conservative 0;  | Mismatches 0;  | Indels 0;   | Gaps 0; |        |                 |
| Qy                    | 1  | GTTGGGGCGGAGCTTACGACTGCAAGAGATAAGTCGTTGCGATCTGTGAAGG   | 60  |         |        |                 |
| Db                    | 243  | GTGGGGTGGGGCTTAAGCATGCAAGAGATAAGTCGTTGCGATCTGTGAAGG  | 302   |         |        |                 |
| Qy                    | 61   | CACGTGTGATTCGATAGGGTCAAGGTCTTCCCTAGTTCCTTCACGCCACG   | 120   |         |        |                 |
| Db                    | 303  | CACGTGTGATTCGATAGGTCAAGGTCTTCCCTAGTTCCTTCACGCCACG  | 362   |         |        |                 |
| Qy                    | 121  | GGAGCTGGCCAGAACCTCTCTGGGTGATGACCTAAGGGATGGTTGTGAGGGCCCT  | 180   |         |        |                 |
| Db                    | 363  | GGAGCTGGCCAGAACCTCTGGGTGATGACCTAAGGGATGGTTGTGAGGGCCCT  | 422   |         |        |                 |
| Or                    | 191  | ATGCTTGACACTGGGATCAGACCTTACCCCTAACGCTTACCCATGAGCTGCT   | 221   |         |        |                 |
| BASE COUNT            | 152 a  | 171 c  | 131 g   | 123 t   |        |                 |
| ORIGIN                |  |  |   |         |        |                 |
| Query Match           | 100.0%   | Score 231; DB 28;  | Length 577;   |         |        |                 |
| Best Local Similarity | 100.0%   | Pred. No. 1..6e-60;  |   |         |        |                 |
| Matches 231:          | Conservative 0;  | Mismatches 0;  | Indels 0;   | Gaps 0; |        |                 |

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapter vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.\*

|                    | BASE COUNT            | 159 a  | 135 g   | 125 t  |
|--------------------|-----------------------|--|---|--|
|                    | ORIGIN                |  |   |  |
| Qy                 | Query Match           | 100.0%   | Score 231;  | DB 28;   |
| Db                 | Best Local Similarity | 100.0%;  | Pred. No. 1   | Length 587;  |
| Qy                 | Matches 231;          | 6e-60;   | Mismatches 0;   | Gaps 0;  |
| Db                 | Qy                    | 1  | GTGGGTGCGAGCTTAAGGACTGACAGGATAGCTGGCATCCCTGGAAAGG 60<br>   <br>61 CACGCTGATTGCAAGGATTCAGTGTCTAGTCCTAGTGTCTAGTCATCC 120<br>   <br>314 CACGCTGATTGCAAGGATTCAGTGTCTAGTCCTAGTGTCTAGTCATCC 255<br>   <br>121 GGAGCTGGCCAAGACCTCTCTGGGTAGAGCTTAAGGGATGGTTTGTA 180<br>   <br>254 GGAGCTGGCCAAGACCTCTCTGGGTAGAGCTTAAGGGATGGTTTGTA 195<br>   <br>181 ATGCCTGCACATGGATCAGACCTCACCTCACCATGAGGCTTGCTT 231<br>   <br>194 ATGCCTGCACATGGATCAGACCTCACCTCACCATGAGGCTTGCTT 144<br> |  |
| Db                 | Qy                    | 1  | GTGGGTGCGAGCTTAAGGACTGACAGGATAGCTGGCATCCCTGGAAAGG 60<br>   <br>246 GTGGGTGCGAGGTTAAGGACTGACAGGATAGCTGGCATCCCTGGAAAGG 187<br>  |  |
| Qy                 | Qy                    | 1  | GTGGGTGCGAGCTTAAGGACTGACAGGATAGCTGGCATCCCTGGAAAGG 60<br>   <br>61 CACGCTGATTGCAAGGATTCAGTGTCTAGTCATCCCTGGAAAGACAGC 120<br>   <br>186 CACGCTGATTGCAAGGATTCAGTGTCTAGTCATCCCTGGAAAGACAGC 127<br>   |  |
| Db                 | Qy                    | 1  | GGAGCTGGCCAAGAACCTCTGGGTAGAGCTTAAGGGATGGTTAGGGCCCT 180<br>   <br>121 GGAGCTGGCCAAGAACCTCTGGGTAGAGCTTAAGGGATGGTTAGGGCCCT 180<br>   <br>126 GGAGCTGGCCAAGAACCTCTGGGTAGAGCTTAAGGGATGGTTAGGGCCCT 67<br>   |  |
| Db                 | Qy                    | 1  | ATGCCTGCACACTGGGATCAGACCTCACCTCACCATGAGGCTTGCTT 231<br>   <br>66 ATGCCTGCACACTGGGATCAGACCTCACCATGAGGCTTGCTT 16<br>  |  |
| RESULT 7           | RESULT 8              | AZ495883/c   | AZ495883  | 627 bp DNA linear GSS 05-OCT-2000  |
| LOCUS              | LOCUS                 | AZ495883/c   | AZ495883  | 627 bp DNA linear GSS 05-OCT-2000  |
| DEFINITION         | DEFINITION            | IM0332803F   | IM0332803F  | Mouse 10kb plasmid UGCM library Mus musculus genomic clone UGCLM0332E03 F, genomic survey sequence.  |
| ACCESSION          | ACCESSION             | AZ495883   | AZ495883  | Accession AZ495883   |
| VERSION            | VERSION               | AZ495883.1   | AZ495883.1  | Version AZ495883.1   |
| KEYWORDS           | KEYWORDS              | GSS.   | GSS.  | Keywords GSS.  |
| SOURCE             | SOURCE                | Mus musculus (house mouse)   | Mus musculus (house mouse)  | Source Mus musculus  |
| ORGANISM           | ORGANISM              | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus.  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus.   | Organism Mus musculus  |
| REFERENCE          | REFERENCE             | Dunn,D., Ayogai,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenah,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weisz,R.  | Dunn,D., Ayogai,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenah,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weisz,R.   | Author Dunn,D., Ayogai,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenah,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weisz,R. |
| AUTHORS            | AUTHORS               | Unpublished  | Unpublished   | Unpublished  |
| ISLAM,H.           | ISLAM,H.              | Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts   | Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  | Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts   |
| 1 (bases 1 to 587) | 1 (bases 1 to 587)    | Contact: Robert B. Weiss   | Contact: Robert B. Weiss  | Contact: Robert B. Weiss   |
| REFERENCE          | COMMENT               | University of Utah Genome Center   | University of Utah Genome Center  | University of Utah Genome Center   |
| AUTHORS            | COMMENT               | Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  | Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT   | University of Utah   |
| ISLAM,H.           | COMMENT               | 84112, USA   | 84112, USA  | Rm. 308 Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT   |
| 1                  | COMMENT               | Phone: 801 585 5606  | Phone: 801 585 5606   | Phone: 801 585 5606  |
| 1                  | COMMENT               | Fax: 801 585 7177  | Fax: 801 585 7177   | Fax: 801 585 7177  |
| 1                  | COMMENT               | Email: ddunn@genetics.utah.edu   | Email: ddunn@genetics.utah.edu  | Email: ddunn@genetics.utah.edu   |
| 1                  | COMMENT               | Insert Length: 10000 Std Error: 0.00   | Insert Length: 10000 Std Error: 0.00  | Insert Length: 10000 Std Error: 0.00   |
| 1                  | COMMENT               | Plate: 0112 row: B column: 13  | Plate: 0112 row: B column: 13   | Plate: 0112 row: E column: 03  |
| 1                  | COMMENT               | Seq primer: CACACAGAAAAAACGCTAATGACC   | Seq primer: CACACAGAAAAAACGCTAATGACC  | Seq primer: CGTTGTAACAGGGCCAGT   |
| 1                  | COMMENT               | Class: Plasmid ends  | Class: Plasmid ends   | Class: Plasmid ends  |
| 1                  | COMMENT               | High quality sequence stop: 587.   | High quality sequence stop: 587.  | High quality sequence stop: 627.   |
| FEATURES           | FEATURES              | Location/Qualifiers  | Location/Qualifiers   | Location/Qualifiers  |
| source             | source                | 1. 587   | 1. 587  | 1. 627   |
|                    |                       | /organism="Mus musculus"   | /organism="Mus musculus"  | /organism="Mus musculus"   |
|                    |                       | /mol_type="genomic DNA"  | /mol_type="genomic DNA"   | /mol_type="genomic DNA"  |
|                    |                       | /strain="C57BL/6J"   | /strain="C57BL/6J"  | /strain="C57BL/6J"   |
|                    |                       | /db_xref="taxon:10090"   | /db_xref="taxon:10090"  | /db_xref="taxon:10090"   |
|                    |                       | /clone="UGCC2M0112B13"   | /clone="UGCC2M0112B13"  | /clone="UGCC2M0112B13"   |
|                    |                       | /sex="Male"  | /sex="Male"   | /sex="Male"  |
|                    |                       | /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"   | /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"  | /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"   |
|                    |                       | /clone_lib="Mouse 10kb plasmid UGCM library"   | /clone_lib="Mouse 10kb plasmid UGCM library"  | /clone_lib="Mouse 10kb plasmid UGCM library"   |
|                    |                       | /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource ( <a href="http://www.jax.org/resources/documents/dnares/">http://www.jax.org/resources/documents/dnares/</a> ). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi 4732114 gb AF129072.1), a copy-number | /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource ( <a href="http://www.jax.org/resources/documents/dnares/">http://www.jax.org/resources/documents/dnares/</a> ). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi 4732114 gb AF129072.1), a copy-number  | /clone_lib="Mouse 10kb plasmid UGCM library"   |

/note="Vector: pMD42nV; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.Jax.org/resources/documents/dnarecs/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (giga32132141/gb!AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance." 190 C 142 g 135 t

| Query   | Match        | Score  | DB         | Length | 627;     |    |
|---------|--------------|--|------------|--------|----------|----|
| Best    | Local        | Similarity   | Pred.      | No.    | 1..e-60; |    |
|         |              |  | Mismatches | 0;     |          |    |
| Matches | Conservative | 0;   | Indels     | 0;     | Gaps     | 0; |
| Y       | 1            | GTGGCTGGAGGCTTAAGGACTGCAAGAGATACTGCTGTTGCATCCTTGAGGG   | 60         |        |          |    |
| b       | 441          | GTGGGTGGAGGCTTAAGGACTGCAAGAGATACTGCTGTTGCATCCTTGAGGG   | 382        |        |          |    |
| y       | 61           | CACCTGTATTGCAATGAAGGTTCAAGTGTCTCAAGTCCCTCCAGAAACGACAG  | 120        |        |          |    |
| b       | 381          | CACGTCTGATTGCAATGAAGGTTCAAGTGTCTCAAGTCCCTCCAGAAACGACAG | 322        |        |          |    |
| y       | 121          | GGAGCTGGCCAAGACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTAGGGCCCT | 180        |        |          |    |
| b       | 321          | GGACCTGGCCAAGACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTAGGGCCCT | 262        |        |          |    |
| y       | 181          | ATGCTGCAACATGGGATCAGACCTCACCTTCACCCATGAGSCTTGT         | 231        |        |          |    |
| b       | 261          | ATGCTGCAACATGGGATCAGACCTCACCTTCACCCATGAGSCTTGT         | 211        |        |          |    |

**RESULT** 9  
 AZ93990                  647 bp          DNA linear  
**ACCSID**                  2M0279D16F          Mouse 10kb Plasmid UGCG2M library Mus musculus genomic  
**DEFINITION**                  clone UGCG2M0279D16 F, genomic survey sequence.  
**ACCESSION**                  AZ93990  
**VERSION**                  AZ993990.1          GI:13865217  
**KEYWORDS**                  GSS.  
**SOURCE**                  Mus musculus (house mouse)  
**ORGANISM**                  Mus musculus  
**MATERIAL**                  Mus musculus

**MATERIALS AND METHODS**  
**REFERENCE** Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** I. (Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meinen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederaussem, A., and Wright, P., Weiss, R.)  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
**Insert Length:** 10000    **Std Dev:** 16  
**Plate:** 0279    **Row:** D    **Column:** 16

| FEATURES  | source  |
|---|---|
| Seq primer: CGTGTAAAAAGACGGCCAG<br>Class: Plasmid ends<br>High quality sequence stop: 647.<br>Location/Qualifiers | 1. 647<br>/organism="Mus musculus"<br>/mol_type="genomic DNA"<br>/strain="C57BL/6J"<br>/db_xref="taxon:10090"<br>/clone="DUGC2N0279D16"<br>/sex="Female"<br>/lab_host="E. coli strain<br>/clone_lib="Mouse 10kb P<br>/note="Vector: PWD42nv";<br>musculus C57BL/6J (female<br>Laboratory Mouse DNA Rese<br>(http://www.jax.org/resource/<br>was hydrodynamically she<br>0.005 inch orifice at co<br>was blunt end-repaired w<br>polynucleotide kinase. A<br>ligated to the blunt end<br>adapted DNA was purified<br>10.5 kb range using prep<br>electrophoresis. Vector<br>of PWD2 (9114732114 [gb]<br>inducible derivative of<br>with adaptors complement<br>purified. The sheared, a<br>adapted vector DNA, an<br>chemically-competent E.<br>and selected for ampicil<br>149 a 146 c 185 g 16 |

| Qy                    | Db   | Qy            | Db        | Qy          | Db | Qy | Db |
|-----------------------|--|---------------|-----------|-------------|----|----|----|
| Query Match           | 100.0%   | Score 231;    | DB 28;    | Length 647; |    |    |    |
| Best Local Similarity | 100.0%   | Pred. No.     | 1.7e-60;  |             |    |    |    |
| Matches 231;          | Conservative 0;  | Mismatches 0; | Indels 0; | Gaps 0;     |    |    |    |
|                       |  |               |           |             |    |    |    |
| 1                     | GTGGGTGCGAGCTAGACTGCAAGGATAGCTTGCTGCATCCGTGGAGG 60                 |               |           |             |    |    |    |
| 235                   | GTGGGTGCGAGCTAGACTGCAAGGATAGCTTGCTGCATCCGTGGAGG 294                |               |           |             |    |    |    |
| 61                    | CACGTCTGATTGCCATGAGGTTCAGTCTCCATTCCCTAGGATAGCTTGCAAGAAAACGCACG 120 |               |           |             |    |    |    |
| 295                   | CACGTCTGATTGCCATGAGGTTCAGTCTCCATTCCCTAGGATAGCTTGCAAGAAAACGCACG 354 |               |           |             |    |    |    |
| 121                   | GGAGCTGGCCAAGAACCTCTCGGGTGTAGCCTAAGGATGGTTTGTGTAGGGCCCT 180        |               |           |             |    |    |    |
| 355                   | GGAGCTGGCCAAGAACCTCTCGGGTGTAGCCTAAGGATGGTTTGTGTAGGGCCCT 414        |               |           |             |    |    |    |
| 181                   | ATGCTTGCACACTGGGATCACCCCTACCTTACCCATAGGCTGTCT 231                  |               |           |             |    |    |    |
| 415                   | ATGCTTGCACACTGGGATCACCCCTACCTTACCCATAGGCTGTCT 465                  |               |           |             |    |    |    |

|           |            |           |          |            |  |                                      |             |        |                         |
|-----------|------------|-----------|----------|------------|--|--------------------------------------|-------------|--------|-------------------------|
| RESULT 10 | CB056516/c | LOCUS     | CB056516 | DEFINITION | NISC_JJ17f07.w1 Soares IMAGE:488516_5,   | 706 bp                               | mRNA        | linear | EST 17-JAN-2003         |
|           |            |           |          |            | 'mRNA sequence.  |                                      |             |        | Mus musculus cDNA clone |
|           |            | ACCESSION | CB056516 | VERSION    | CB056516.1   | EST.                                 | GI:27794803 |        |                         |
|           |            | KEYWORDS  |          | SOURCE     | Mus musculus (house mouse)   |                                      |             |        |                         |
|           |            | ORGANISM  |          |            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurogathi; Muridae; Murinae; Mus. 1 (bases 1 to 706) |                                      |             |        |                         |
|           |            | REFERENCE |          | AUTHORS    | NCI-CGAP   | http://www.ncbi.nlm.nih.gov/ncicgap. |             |        |                         |
|           |            | TITLE     |          |            | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  |                                      |             |        |                         |
|           |            | JOURNAL   |          |            | Unpublished  |                                      |             |        |                         |

| COMMENT     | Contact: Robert Strausberg, Ph.D.<br>Email: cgabps-r@mail.nih.gov<br>cDNA Library Preparation:<br>DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)<br>Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: InfoImage.lnl.gov<br>Plate: LLM10799 row: K column: 13<br>Seq primer: T7 primer,<br>Location/Qualifiers   |
|-------------|---|
| FEATURES    | source<br>1. 706      /organism="Mus musculus"<br>/mol_type="mRNA"<br>/db_xref="taxon:10090"<br>/clone="IMAGE:4848816"<br>/tissue_type="pituitary gland"<br>/dev_stage="juvenile, 13-15 days"<br>/lab_host="dh10B (phage-resistant)"<br>/clone_id="Soares NMNP13-15."<br>/notes="Organ: brain; Vector: pT7T3D-Paci; Site: 1; NotI: Site 2; EcoRI: 1st strand cDNA was primed with a NotI-0190 (dR) primer<br>5'-AACGTGAGATTCGGCCCGCTGTACCGATGTTTTTTTTTTTT-3'; double-stranded cDNA was ligated to EcoRI adaptors 5'-AATTCGCACTGG-3' and 5'-CCCTGCCTCG-3' (Pharmacia), digested with NotI and cloned into the NotI and EcoRI sites of the pT7T3D-Paci vector. Library went through one round of normalization, and was constructed in the laboratory of M. Bento Soares (University of Iowa)." |
| BASE COUNT  | 157 a 199 c 172 g 178 t   |
| ORIGIN      |   |
| Query Match | 100.0% Score 231; DB 14; Length 706;<br>Best Local Similarity 100.0%; Pred No. 1.8e-60;<br>Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;   |
| QY          | 1 GTGGGGTGGCAAGGGCTAAGGTTCAGGTCTAGTTCCCTCCCCAGAAAAACGACAGC 60   |
| Db          | 634 GTGGGGTGGCAAGGGCTAAGGTTCAGGTCTAGTTCCCTCCCCAGAAAAACGACAGC 575  |
| QY          | 61 CACGTCGATTGCAAGACCTCTCTGGGTATGAGCCTTAAGGTGGTTTGTCAGT 120   |
| Db          | 574 CACGTCGATTGCAAGACCTCTCTGGGTATGAGCCTTAAGGTGGTTTGTCAGT 515  |
| QY          | 121 GGAGCTGGCCAAGAAAGCTCTCTGGGTATGAGCCTTAAGGTGGTTTGTCAGT 180  |
| Db          | 514 GGAGCTGGCCAAGAAAGCTCTCTGGGTATGAGCCTTAAGGTGGTTTGTCAGT 455  |
| QY          | 181 ATGCTTGCACACTGGGATCAAGACCTCTACCTTACCCATGAGCTGCTT 231  |
| Db          | 454 ATGCTTGCACACTGGGATCAAGACCTCTACCTTACCCATGAGCTGCTT 404  |
| RESULT 11   | A2994818/c A2994818/c   |
| LOCUS       | A2994818/c 721 bp DNA linear  |
| DEFINITION  | 2M028024F Mouse 10kb plasmid unGC2M library Mus musculus genomic clone UGGC2A028024 F, genomic survey sequence.   |
| ACCESSION   | AZ994818  |
| VERSION     | AZ994818.1 GI:13866045  |
| KEYWORDS    | GSS.  |
| SOURCE      | Mus musculus (house mouse)  |
| ORGANISM    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |
| REFERENCE   | 1 (bases 1 to 721)  |
| AUTHORS     | Dunn,D., Aoyagi,A., Barber,M., Beacons,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmud,M., Meinen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.   |
| TITLE       | Mouse whole genome scaffolding with paired end reads from 10kb  |



293 CATGTCGATTGCATGGAGGTTCAAGTGTCCCTAGTGTCCCCTCCCCCAGGAAAAACGACACG 352  
 constructed by Xiaohong Wang and Yulan Piao.<sup>\*</sup>

BASE COUNT 126 a 138 c 131 g 123 t  
 ORIGIN

Query Match 99.3%; Score 229.4; DB 9;  
 Best Local Similarity 99.6%; Pred. No. 4.9e-60;  
 Matches 230; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

QY 1 GTGGGGTGCAGGTAAGGACTGACAGGGATAAGGATAGCTGTCATCCTGTGGAAAG 60  
 Db 465 GTGGGGTGCAGGTAAGGACTGACAGGGATAAGGATAGCTGTCATCCTGTGGAAAG 406

Db 61 CACGCCTGATTGCGATGAAGGTTCAAGTGTCCCTCCCCAGGAAAACGACACG 120  
 Qy 405 CACGCCTGATTGCGATGAAGGTTCAAGTGTCCCTCCCCAGGAAAACGACACG 316

Db 121 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 180  
 Qy 121 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 180

Db 345 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 286

Db 181 ATGGTTGCCACACTGGGACAGACTCTACCTCATGGCTTGCTT 231  
 Qy 285 ATGGCTGCCACACTGGGACAGACTCTACCTCATGGCTTGCTT 235

---

RESULT 14  
 LOCUS AW558986  
 DEFINITION L0302F10-3 NIA Mouse 518 bp mRNA linear EST 31-AUG-2000  
 ACCESSION AW558986  
 VERSION GI:7204415  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Tanaka,T.S., Jarad,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac  
 M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.  
 III, Becker,K.G. and Ko,M.S.H.  
 TITLE Genome-wide expression profiling of mid-gestation placenta and  
 embryo using a 15,000 mouse developmental cDNA microarray  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)  
 MEDLINE 10922068  
 PUBMED  
 COMMENT Contact: George J. Kargul  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: dna@igsun.grc.nia.nih.gov  
 Plate: L0302F10 row: F column: 10  
 Seq primer: -2LM13 Forward  
 High quality sequence Stop: 518  
 POLYA=Yes.

FEATURES source  
 1. .518  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="naEST:10302F10-3"  
 /clone\_xref="Taxon:10090"  
 /sex="Female"  
 /dev\_stage="Newborn Ovary"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1 (Gibco/BRU Life Technology);  
 Site\_1: Sali; Site\_2: NotI; Total RNAs were extracted from  
 7 Newborn ovary. The double-stranded cDNA was synthesized  
 by Gibco's kit with an Oligo(dT) primer [NotI  
 primer adapter from Gibco/BRL]  
 [5'-pGCTAGTCTAGATCCGAGCGCCCTTTTTTTTT-3'] from  
 2.56ug of total RNA. The double-stranded cDNAs were  
 treated with T4 DNA polymerase and purified by  
 ethanol precipitation. The cDNAs were ligated to  
 Lone-linker LL-Sal3 (include SalI sequence). The cDNAs  
 were purified by Centrifuge 100. Then, cDNAs were amplified  
 by long range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centrifuge 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into Sall/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was

BASE COUNT 126 a 138 c 131 g 123 t  
 ORIGIN

Query Match 99.3%; Score 229.4; DB 9;  
 Best Local Similarity 99.6%; Pred. No. 4.9e-60;  
 Matches 230; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

QY 1 GTGGGGTGCAGGTAAGGACTGACAGGGATAAGGATAGCTGTCATCCTGTGGAAAG 60  
 Db 465 GTGGGGTGCAGGTAAGGACTGACAGGGATAAGGATAGCTGTCATCCTGTGGAAAG 406

Db 61 CACGCCTGATTGCGATGAAGGTTCAAGTGTCCCTCCCCAGGAAAACGACACG 120  
 Qy 405 CACGCCTGATTGCGATGAAGGTTCAAGTGTCCCTCCCCAGGAAAACGACACG 316

Db 121 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 180  
 Qy 121 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 180

Db 345 GGAGCTGGCCAAGAACCTCTGGGTGATGGCTTGTGTAAGGGCCCT 286

Db 181 ATGGTTGCCACACTGGGACAGACTCTACCTCATGGCTTGCTT 231  
 Qy 285 ATGGCTGCCACACTGGGACAGACTCTACCTCATGGCTTGCTT 235

RESULT 15  
 LOCUS BM219777/C  
 DEFINITION C0929H06-3 NIA Mouse 520 bp mRNA linear EST 07-JUN-2003  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 3', mRNA sequence.

ACCESSION BM219777  
 VERSION BM219777.2  
 SOURCE EST.  
 SOURCE Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 520)

REFERENCE  
 AUTHORS Piao,Y., Ko,N.T., Lin,M.K. and Ko,M.S.H.  
 TITLE Construction of long-transcript enriched cDNA libraries from  
 submicrogram amounts of total RNAs by a universal PCR amplification  
 method  
 JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
 MEDLINE 21429098  
 PUBLMED 11544199  
 COMMENT On Dec 14, 2001 this sequence version replaced gi:17779702.  
 Contact: Dawood B. Dukekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@gsun.grc.nia.nih.gov  
 Plate: C0929 row: H column: 06  
 Seq primer: -2LM13 Forward  
 High quality sequence stop: 520  
 POLYA=Yes.

FEATURES source  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="naEST:10302F10-3"  
 /clone\_xref="Taxon:10090"  
 /sex="Male"  
 /dev\_stage="Newborn Ovary"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1 (Gibco/BRU Life Technology);  
 Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from  
 7 Newborn ovary. The double-stranded cDNA was synthesized  
 by Gibco's kit with an Oligo(dT) primer [NotI  
 primer adapter from Gibco/BRL]  
 [5'-pGCTAGTCTAGATCCGAGCGCCCTTTTTTT-3'] from  
 2.56ug of total RNA. The double-stranded cDNAs were  
 treated with T4 DNA polymerase and purified by  
 ethanol precipitation. Then, the cDNAs were ligated to  
 Lone-linker LL-Sal3 (include SalI sequence). The cDNAs  
 were purified by Centrifuge 100. Then, cDNAs were amplified  
 by long range high fidelity PCR using Takara's Ex Taq  
 polymerase. Then, the cDNAs were purified by  
 phenol/chloroform and by Centrifuge 100. The cDNAs were  
 digested with SalI and NotI enzymes. Then, the cDNAs were  
 size selected by Gibco's Size Fractionation Column. The  
 cDNAs were cloned into Sall/NotI site of pSPORT1 plasmid  
 vector. The DH10B E. coli host was transformed with the  
 ligation mixture by chemical method. The library was

ORGANISM="Mus musculus"  
 mol\_type="mRNA"  
 strain="C57BL/6J"  
 db\_xref="naEST:C0929H06-3"  
 dev\_stage="1.5-dpc"  
 lab\_host="DH10B"  
 clone\_lib="NIA Mouse 12.5-dpc Male Genital  
 Ridge/Mesonephros cDNA Library (Long)"  
 note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2:  
 NotI; Mouse cDNA project by the Laboratory of Genetics,  
 National Institute on Aging (NIA), Intramural Research

Program, NIH (<http://igsun.grc.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [In vitro]:  
 5'-GACTAGTCTATCGGAGGCCGCCTTTTTTTTTT-3'] from 1.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centrifugation 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and centrifugation 100. The cDNAs were digested with Sall and NotI enzymes.

and cloned into SalI/NotI site of pSPORII plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yuan Piao (NIA).<sup>1</sup>

| Query             | Match   | Score | DB 12;     | Length   |
|-------------------|---|-------|------------|----------|
| Best <sub>i</sub> | Local Similarity  | Score | DB 12;     | Length   |
| Matches           | Matches   | Pred. | No.        | 4.9e-60; |
| Dy                | Dy  | 0;    | Mismatches | Indels   |
| Dy                | Dy  | 0;    | Gaps       | 0;       |
| Dy                | Dy  | 0;    |            |          |
| 1                 | GTGGGTGCGAGGCCATTAGCACTGCAGAGGATAGCTGCTGTCATCCTGTGAAAGG       | 60    |            |          |
| 465               | GTGGGTGCGAGGCCATTAGCACTGCAGAGGATAGCTGCTGTCATCCTGTGAAAGG       | 406   |            |          |
| 61                | CACGTCUGATTCATGAAGGTTCACTGGGTGATGCTCCCTCCAGAAAACGCACG         | 120   |            |          |
| 405               | CACGTCGTGATGCCATGAAGGTTCACTGGGTGATGCTCCCTCCAGAAAACGCACG       | 346   |            |          |
| 121               | GGAGCTGGCCCAAAGACCCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTGTAGGGCCCT | 180   |            |          |
| 345               | GGAGCTGGCCCAAAGACCCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTGTAGGGCCCT | 286   |            |          |
| 181               | ATGGCTTGACACTGGGATCAGACUCCUACCTTACCCATGAGGTGTGCTT             | 231   |            |          |
| 285               | ATGCTTGACACTGGGATCAGACUCCUACCTTACCCATGAGGTGTGCTT              | 235   |            |          |

Search completed: October 9, 2003, 22:40:58  
Total time: 2007-2008

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 9, 2003, 22:44:04 ; Search time 206 Seconds  
(without alignments)  
2909.716 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtgggtgcgaggctaa...ttcacccataggcttgctt 231

Scoring table: IDENTITY\_NUC  
GapOp 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:  
 1: /cgn2\_6\_ptodata/1/pubpna/US07\_PUBCOMB.seq:  
 2: /cgn2\_6\_ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
 3: /cgn2\_6\_ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
 4: /cgn2\_6\_ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
 5: /cgn2\_6\_ptodata/1/pubpna/PUBCOMB.seq:  
 6: /cgn2\_6\_ptodata/1/pubpna/PCRTS\_PUBCOMB.seq:  
 7: /cgn2\_6\_ptodata/1/pubpna/US08\_PUB.seq:  
 8: /cgn2\_6\_ptodata/1/pubpna/PUBCOMB.seq:  
 9: /cgn2\_6\_ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
 10: /cgn2\_6\_ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
 11: /cgn2\_6\_ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
 12: /cgn2\_6\_ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
 13: /cgn2\_6\_ptodata/1/pubpna/US08\_PUBCOMB.seq:  
 14: /cgn2\_6\_ptodata/1/pubpna/US10\_PUBCOMB.seq:  
 15: /cgn2\_6\_ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
 16: /cgn2\_5\_ptodata/1/pubpna/US60\_PUBCOMB.seq:  
 17: /cgn2\_6\_ptodata/1/pubpna/US60\_PUBCOMB.seq:  
 \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Length | DB ID  | Description                         |
|------------|-------|--------------|--------|-------------------------------------|
| C 1        | 214.2 | 92.7         | 185548 | Sequence 62, Appl                   |
| C 2        | 92.6  | 40.1         | 432    | Sequence 502, Appl                  |
| C 3        | 89.4  | 38.7         | 1913   | Sequence 40, Appl                   |
| C 4        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 5        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 6        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 7        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 8        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 9        | 86.2  | 37.3         | 1434   | Sequence 13, Appl                   |
| C 10       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 11       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 12       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 13       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 14       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 15       | 81.4  | 35.2         | 1400   | Sequence 14, Appl                   |
| C 16       | 70.4  | 30.5         | 251364 | Sequence 61, Appl                   |
| C 17       | 70.4  | 30.5         | 251364 | Sequence 79, Appl                   |
| C 18       | 70.4  | 34           | 14.7   | Sequence 36, Appl                   |
| C 19       | 70.4  | 30.4         | 13.2   | Sequence 475, Appl                  |
| C 20       | 30.4  | 30           | 13.0   | Sequence 447, Appl                  |
| C 21       | 30    | 30           | 13.0   | Sequence 447, Appl                  |
| C 22       | 30    | 30           | 13.0   | Sequence 447, Appl                  |
| C 23       | 30    | 30           | 13.0   | Sequence 457, Appl                  |
| C 24       | 30    | 30           | 13.0   | Sequence 609, Appl                  |
| C 25       | 29.6  | 29.6         | 12.8   | Sequence 668, Appl                  |
| C 26       | 29.6  | 29.6         | 12.8   | Sequence 968, Appl                  |
| C 27       | 29.6  | 29.4         | 12.7   | Sequence 1049, Appl                 |
| C 28       | 29.4  | 29.4         | 12.7   | Sequence 12917, Appl                |
| C 29       | 29.4  | 12.7         | 32.9   | Sequence 14097, Appl                |
| C 30       | 29.2  | 29.2         | 12.6   | Sequence 540, Appl                  |
| C 31       | 29.2  | 29.2         | 12.6   | Sequence 2977, Appl                 |
| C 32       | 29.2  | 29.2         | 12.6   | Sequence 2977, Appl                 |
| C 33       | 28.6  | 28.6         | 12.4   | Sequence 360, Appl                  |
| C 34       | 28.6  | 28.6         | 12.4   | Sequence 427, Appl                  |
| C 35       | 28.6  | 28.6         | 12.4   | Sequence 427, Appl                  |
| C 36       | 28.6  | 28.6         | 12.4   | Sequence 564, Appl                  |
| C 37       | 28.6  | 28.6         | 12.4   | Sequence 564, Appl                  |
| C 38       | 28.6  | 28.6         | 12.4   | Sequence 564, Appl                  |
| C 39       | 28.6  | 28.6         | 12.4   | Sequence 564, Appl                  |
| C 40       | 28.6  | 28.6         | 12.4   | Sequence 1197, Appl                 |
| C 41       | 28.6  | 12.4         | 6549   | Sequence 10-0-240-965-45, Appl      |
| C 42       | 28.4  | 28.4         | 12.3   | Sequence 2449, Appl                 |
| C 43       | 28.4  | 28.4         | 12.3   | Sequence 2449, Appl                 |
| C 44       | 28.4  | 12.3         | 10472  | Sequence 10-0-27-632-112101, Appl   |
| C 45       | 28.4  | 12.3         | 10472  | Sequence 10-0-09-764-841-1615, Appl |
| C 46       | 28.4  | 12.3         | 10472  | Sequence 10-0-09-764-841-1615, Appl |
| C 47       | 28.4  | 12.3         | 10472  | Sequence 10-0-09-764-841-1615, Appl |

## ALIGNMENTS

RESULT 1  
US-10-115-523-62/c  
; Sequence 62, Application US/10175523  
; Publication No. US20030036264A1  
; GENERAL INFORMATION:  
; APPLICANT: Brockman, Jeffrey  
; APPLICANT: Evans, David  
; APPLICANT: Hook, Derek  
; APPLICANT: Klimczak, Leszek  
; APPLICANT: Laeng, Pascal  
; APPLICANT: Palfreyman, Michael  
; APPLICANT: Rajan, Prithi  
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHRS)  
; FILE REFERENCE: 3235/1J795-US3  
; CURRENT APPLICATION NUMBER: US/10/175,523  
; PRIOR APPLICATION NUMBER: US 60/299,151  
; PRIORITY FILING DATE: 2002-06-18  
; PRIORITY FILING DATE: 2001-09-07  
; PRIORITY FILING DATE: 2001-09-25  
; PRIORITY FILING DATE: 2001-11-14  
; PRIORITY FILING DATE: 2002-01-18  
; PRIORITY FILING DATE: 2002-03-04  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO: 62  
; LENGTH: 185548  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-175-523-62

Query Match Score 92.7%; Best Local Similarity 98.3%; Matches 227; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query 1 GGGGCTGGAGGCTAAAGCACTGGCACAGGATAGCTGTGTCATCCTGTTGAAGG 60  
 Database 52312 GTGGGTGGGGTAAGCACTGCAGGATACTGCTCTGGCAAGG 52253

Query 61 CACGCTGATGTCATGAGGTTCACTGTCCTAGTCCCTCCCCAGAAAAGACAGC 120  
 Database 52252 TATGTCGA-TCCATGAAGGTTCACTGCTCTAGTCCCTCCCCAGAAAAGACAGC 52194

Query 121 GGAGCTGCAAGAACCTCTGGGATGAGCCTAAGGGATGGTTTGTTGTTGTT 180  
 Database 52193 GGAGCTGCAAGAACCTCTGGGATGAGCCTAAGGGATGGTTTGTTGTTGTT 52134

Query 181 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCATGAGCTTGCTT 231  
 Database 52133 ATGCTTGACACTGGGATCAGACCTTCACCATGAGCTTGCTT 52083

---

**RESULT 2**  
 Sequence 502, Application US/09933797  
 Patent No. US20020155119A1

GENERAL INFORMATION:  
 APPLICANT: Robert A. Sikes et al.  
 TITLE OF INVENTION: Isolation and Use of Fetal Urogenital Sinus Expressed Sequences  
 FILE REFERENCE: 9901-007-999

CURRENT APPLICATION NUMBER: US/09/933,797  
 CURRENT FILING DATE: 2001-08-22  
 PRIOR APPLICATION NUMBER: US/09/482,933  
 PRIOR FILING DATE: 2000-01-14  
 PRIOR APPLICATION NUMBER: PCT/US99/10746  
 PRIOR FILING DATE: 1999-05-14  
 PRIOR APPLICATION NUMBER: 60/085,383  
 PRIOR FILING DATE: 1998-01-14  
 NUMBER OF SEQ ID NOS: 811  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO: 502  
 LENGTH: 432  
 TYPE: DNA  
 ORGANISM: Murine  
 US-09-933-716-1.rnpb

Query Match 40.1%; Score 92.6; DB 10; Length 432;  
 Best Local Similarity 76.0%; Pred. No. 9\_9e-24;  
 Matches 130; Conservative 0; Mismatches 34; Indels 7; Gaps 1;

Query 55 GGAAGGCACTGCTGATTGCAAGGTTCACTGTCCTCCCCAGAAAAC 114  
 Database 416 GGAGACATGTCATCTCATGAGGTCTAGTGCCTCTAGTCCCTCCCCAGAAAAC 357

Query 115 GACAGGGAGCTGGCAAGACCTCTGGTCACTGCTTGGGATGGTTT 167  
 Database 356 GACAGGGAGCTGGCAAGACCTCTGGTCACTGCTTGGGATGGTTT 297

Query 168 GTGTAAGGCCCTATGCTGACACTGGGATCAGACCTCTACCTTCACCC 218  
 Database 296 GTACATGGCTCTTACACAGCTGGGATGACCTCTACCTTCACCC 246

---

**RESULT 4**  
 Sequence 13, Application US/09799462A  
 Patent No. US20020160970A1

GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Grulla Szalay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Shuman White & McAuliffe  
 STREET: 4250 Executive Square, 7th Floor  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/799,462A  
 FILING DATE: 10-SEP-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/035,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/095,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/082,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/029,822  
 FILING DATE: 10-APR-1996

ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L  
 REGISTRATION NUMBER: 33-779  
 TELECOMMUNICATION INFORMATION:  
 REFERENCE/DOCKET NUMBER: 24601-402G  
 TELEPHONE: 858-587-5360  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 HYPOTHETICAL: NO

---

**RESULT 3**  
 US-10-225-810-40/C  
 Sequence 40, Application US/10225810  
 Publication No. US200301512A1

GENERAL INFORMATION:  
 APPLICANT: Birmingham, Jr., John R.  
 TITLE OF INVENTION: Tramdorin and Methods of Using Tramdorin

FILE REFERENCE: McLaugh-07165  
 CURRENT APPLICATION NUMBER: US/10/225,810  
 CURRENT FILING DATE: 2002-08-21  
 NUMBER OF SEQ ID NOS: 76  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO: 40

ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE:  
 SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-09-799-462A-13

Query Match 37.3%; Score 86.2; DB 10; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

Qy 55 GGAGGCCACGTCTGATGTCATGAGGGTCAGTGTCCATTAGTTCCCTTCCCAGGAAAC 114  
 Db 900 GGGGAATGATGTCATCTCATGAGGTTCAGTGTCCCTTCCCAGGCAAAAC 959

Qy 115 GACACGGGAGCAGGGCCAGAACCTCTCGGGTGA-----TGAGCCTTAAGGGATGGTT 167  
 Db 950 GACACGGGAGCAGGTCAAGGGTCTCGGGTAAAGGCTGAGCCTGGAGCTAAATCC 1019

Qy 168 GTGTTAGGCCCTATGCTTGACACTGGGATCAGACCTTACCTGACTGACCC 218  
 Db 1020 GTACATGGCTCCCTTAACCTACACAGGGATTTGACCTCTACCTCACT 1070

RESULT 5  
 US-09-936-911A-13  
 Sequence 13, Application US/09836911A  
 Publication No. US2003003361A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 Spalay, Aladar  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Ehrman White & McAuliffe  
 STREET: 4350 La Jolla Village Drive, 6th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Disquette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/09836, 911A  
 FILING DATE: 17-Apr-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/635, 682  
 FILING DATE: 10-Apr-1997  
 APPLICATION NUMBER: 08/695, 191  
 FILING DATE: 07-Aug-1996  
 APPLICATION NUMBER: 08/682, 080  
 FILING DATE: 15-Jul-1996  
 APPLICATION NUMBER: 08/629, 822  
 FILING DATE: 10-Apr-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402I  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEX: 858-587-5360  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA

RESULT 6  
 US-10-125-767-13  
 Sequence 13, Application US/10125767  
 Publication No. US2002160410A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
 METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
 STREET: 4350 La Jolla Village Drive, 7th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/125, 767  
 FILING DATE: 17-Apr-2002  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 09/724, 693  
 FILING DATE: 28-Nov-2000  
 APPLICATION NUMBER: 08/835, 682  
 FILING DATE: 10-Apr-1997  
 APPLICATION NUMBER: 08/695, 191  
 FILING DATE: 07-Aug-1996  
 APPLICATION NUMBER: 08/682, 080  
 FILING DATE: 15-Jul-1996  
 APPLICATION NUMBER: 08/629, 822  
 FILING DATE: 10-Apr-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402J  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEX: 858-587-5360  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE: SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-09-836-911A-13

HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE: SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-09-836-911A-13

Query Match 37.3%; Score 86.2; DB 11; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

Qy 55 GGAGGCCACGTCTGATGTCATGAGGGTCAGTGTCCATTAGTTCCCTTCCCAGGAAAC 114  
 Db 900 GGGGAATGATGTCATCTCATGAGGTTCAGTGTCCCTTCCCAGGCAAAAC 959

Qy 115 GACACGGGAGCAGGGCCAGAACCTCTCGGGTGA-----TGAGCCTTAAGGGATGGTT 167  
 Db 950 GACACGGGAGCAGGTCAAGGGTCTCGGGTAAAGGCTGAGCCTGGAGCTAAATCC 1019

Qy 168 GTGTTAGGCCCTATGCTTGACACTGGGATCAGACCTTACCTGACTGACCC 218  
 Db 1020 GTACATGGCTCCCTTAACCTACACAGGGATTTGACCTCTACCTCACT 1070

RESULT 5  
 US-09-936-911A-13  
 Sequence 13, Application US/09836911A  
 Publication No. US2003003361A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 Spalay, Aladar  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Ehrman White & McAuliffe  
 STREET: 4350 La Jolla Village Drive, 6th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Disquette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/09836, 911A  
 FILING DATE: 17-Apr-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/635, 682  
 FILING DATE: 10-Apr-1997  
 APPLICATION NUMBER: 08/695, 191  
 FILING DATE: 07-Aug-1996  
 APPLICATION NUMBER: 08/682, 080  
 FILING DATE: 15-Jul-1996  
 APPLICATION NUMBER: 08/629, 822  
 FILING DATE: 10-Apr-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402I  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEX: 858-587-5360  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA

RESULT 6  
 US-10-125-767-13  
 Sequence 13, Application US/10125767  
 Publication No. US2002160410A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND  
 METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
 STREET: 4350 La Jolla Village Drive, 7th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/125, 767  
 FILING DATE: 17-Apr-2002  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 09/724, 693  
 FILING DATE: 28-Nov-2000  
 APPLICATION NUMBER: 08/835, 682  
 FILING DATE: 10-Apr-1997  
 APPLICATION NUMBER: 08/695, 191  
 FILING DATE: 07-Aug-1996  
 APPLICATION NUMBER: 08/682, 080  
 FILING DATE: 15-Jul-1996  
 APPLICATION NUMBER: 08/629, 822  
 FILING DATE: 10-Apr-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402J  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEX: 858-587-5360  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA

TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE:  
 SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-10-1257-767-13

Query Match Score 86.2; DB 13; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
 Sequence Description: SEQ ID NO: 13:  
 US-10-151-081-13

Qy 55 GGAAGGCCACGTCTGATGCCATTGAAAGTTCAAGTCAGTGCCTTCAGTCAGTCCCTTCAGGCCAGAAAAC 114  
 Db 900 GGGGACATTCATTTTCATGAGGTTAGTGTCTAGTCAGTCAGTCCCTTCAGGCCAAAC 959

Qy 115 GACACGGGAGCTGGCCAGAACCTCTGGGTGA----TGAGCCTTAAGGGATGGTTT 167  
 Db 960 GACACGGGAGCCAGTCAGGCTGGTGTGGTAAAGCCCTGGAGCTTAATCCT 1019

Qy 168 GTTAGAGGGCCCTATGGCTGACACTGGGATCACACCTACCC 218  
 Db 1020 GTACATGGCTCCTTACCTACACACTGGGATTGACCTCTATCCACTC 1070

RESULT 7  
 US-10-151-081-13  
 Sequence 13, Application US/10151081  
 Publication No. US20030083293A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadjacyzky, Gyula  
 SZALAY, Aladar  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF  
 AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller, Ehrman White & McAuliffe  
 STREET: 4250 Executive Square, 7th Floor  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/151,081  
 FILING DATE: 16-May-2002  
 PRIORITY APPLICATION DATA:  
 CLASSIFICATION: <Unknown>  
 APPLICATION NUMBER: 09/799,462  
 FILING DATE: HEREWITH 05-MAR-2001  
 APPLICATION NUMBER: 09/724,633  
 FILING DATE: HEREWITH 28-NOV-2000  
 APPLICATION NUMBER: 08/635,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephen L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402L  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-587-5360  
 TELEFAX: 858-587-5360  
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE:  
 SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-10-151-081-13

Query Match Score 86.2; DB 14; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 3.1e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;  
 Sequence Description: SEQ ID NO: 13:  
 US-10-151-081-13

Qy 55 GGAAGGCCACGTCTGATGCCATTGAAAGTTCAAGTCAGTGCCTTCAGTCAGTCCCTTCAGGCCAGAAAAC 114  
 Db 900 GGGGACATTCATTTTCATGAGGTTAGTGTCTAGTCAGTCAGTCCCTTCAGGCCAAAC 959

Qy 115 GACACGGGAGCTGGCCAGAACCTCTGGGTGA----TGAGCCTTAAGGGATGGTTT 167  
 Db 960 GACACGGGAGCCAGTCAGGCTGGTGTGGTAAAGCCCTGGAGCTTAATCCT 1019

Qy 115 GACACGGGAGCTGGCCAGAACCTCTGGGTGA----TGAGCCTTAAGGGATGGTTT 167  
 Db 960 GACACGGGAGCCAGTCAGGCTGGTGTGGTAAAGCCCTGGAGCTTAATCCT 1019

Qy 168 GTTAGAGGGCCCTATGGCTGACACTGGGATTGACCTCTATCCACTC 218  
 Db 1020 GTACATGGCTCCTTACCTACACACTGGGATTGACCTCTATCCACTC 1070

RESULT 8  
 US-10-287-313-13  
 Sequence 13, Application US/10287313  
 Publication No. US20030101480A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadjacyzky, Gyula  
 SZALAY, Aladar  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller, Ehrman White & McAuliffe  
 STREET: 4250 Executive Square, 7th Floor  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/151,081  
 FILING DATE: 16-May-2002  
 PRIORITY APPLICATION DATA:  
 CLASSIFICATION: <Unknown>  
 APPLICATION NUMBER: 09/799,462  
 FILING DATE: HEREWITH 05-MAR-2001  
 APPLICATION NUMBER: 09/724,633  
 FILING DATE: HEREWITH 28-NOV-2000  
 APPLICATION NUMBER: 08/635,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephen L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402L  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-587-5360  
 TELEFAX: 858-587-5360  
 TELEX: <Unknown>

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/287,313  
 FILING DATE: 01-NOV-2000  
 APPLICATION NUMBER: US20030101480A1-2002  
 CLASSIFICATION: <Unknown>  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 09/724,726  
 FILING DATE: 28-NOV-2000  
 APPLICATION NUMBER: 08/835,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephen L  
 REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24601-402L  
 TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: 24601-402M  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEFAX: 858-587-5360  
 TELERX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE:  
 SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 US-10-287-313-13:

Query Match 37.3%; Score 86.2; DB 14; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 3.1e-21; Length 1434;  
 Matches 128; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

Qy 55 GGAGGGAGCTCATGATGAGGTCAAGCTTCCCTTCCCAGAAAC 114  
 Db 900 GGGGACAGTCATCTTCATGAGGTCAAGCTTCTGGTGA ---TGACCCTAAGGGATGGTT 167  
 Qy 115 GACACGGAGCTGGCCAAGACCTCTGGTCAAGCTTCAGTGTCTTAGTGTCTGGTAAAGCTGTGAGCTTAACCTC 1019  
 Db 960 GACACGGAGCTGGCCAAGACCTCTGGTCAAGCTTCAGTGTCTGGTAAAGCTGTGAGCTTAACCTC 1019  
 Qy 168 GTCTAGGGCCCTATGCTTCACACTGGGATCAACACCCTACTTCACCC 218  
 Db 1020 GTACATGGCTCTTACCTACACATGGGATTGACCTCTATCTCACTC 1070

RESULT 9  
 Sequence 13, Application US/10219694  
 Publication No. US20030108914A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOME, USES THEREOF AND METHODS  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Heller Ehrman White & McAuliffe LLP  
 STREET: 4350 La Jolla Village Drive, 7th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/219,694  
 FILING DATE: 14-Aug-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: <Unknown>  
 FILING DATE: 28-NOV-2000  
 APPLICATION NUMBER: 08/835,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L.  
 REGISTRATION NUMBER: 33,779

RESULT 10  
 Sequence 14, Application US/09799462A  
 Publication No. US20020160970A1  
 GENERAL INFORMATION:  
 APPLICANT: Hadlaczky, Gyula  
 SZALAY, Aladar  
 TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller Ehrman White & McAuliffe  
 STREET: 4250 Executive Square, 7th Floor  
 CITY: La Jolla  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/799,462A  
 FILING DATE: 10-Sep-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/835,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seidman, Stephanie L.  
 REGISTRATION NUMBER: 33,779

REGISTRATION NUMBER: 33,779  
 REFERENCE/DOCKET NUMBER: 24,601-4-02G  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 858-450-8403  
 TELEFAX: 858-587-5360  
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1400 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 HYPOPOLETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: <Unknown>  
 ORIGINAL SOURCE: SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
 US-09-799-462A-14

Query Match 35.2%; Score 81.4; DB 10; Length 1400;  
 Best Local Similarity 71.9%; Pred. No. 1.7e-19;  
 Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

|    |     |  |     |
|----|-----|--|-----|
| Qy | 55  | GGAAGGCACCTCTGATTCATGAACTGAAAGTCAGTGTCCCTCAGGAAAAAC    | 114 |
| Db | 771 | GGGAGACATGNCATTTCAAGAAGGTGAGTTGACTCTCCAGGCCAAC         | 830 |
| Qy | 115 | GACACGGAGCTGGCCAAGACCTCTCTGGTGA-----TGAGCTTAAGGGATGGTT | 167 |
| Db | 831 | GACACGGAGCTGGTGTGGTAAAGCCTGAGCTTAAGGGCTAATCC           | 890 |
| Qy | 168 | GTGTAGGGCCCTATGCTTCACACTGGGATCAGACCTACCTTCACC          | 218 |
| Db | 891 | GTACATGGCTCTTACATCACACTGGGATTCACACTGGGATTCACCTC        | 941 |

RESULT 11  
 US-09-836-911A-14  
 Sequence 14, Application US/09836911A  
 Publication No. US20030033517A1

GENERAL INFORMATION:  
 APPLICANT: Hadjaciicky, Gyula Szalay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Heller, Ehrman White & McAuliffe  
 STREET: 4350 La Jolla Village Drive, 6th Floor  
 CITY: San Diego  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92122

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/836,911A  
 FILING DATE: 17-Apr-2002  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/835,682  
 FILING DATE: 10-APR-1997  
 APPLICATION NUMBER: 08/695,191  
 FILING DATE: 07-AUG-1996  
 APPLICATION NUMBER: 08/682,080  
 FILING DATE: 15-JUL-1996  
 APPLICATION NUMBER: 08/629,822  
 FILING DATE: 10-APR-1996  
 ATTORNEY/AGENT INFORMATION:

FILING DATE: 10-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24601-402J  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-125-767-14

| Query Match | Score | DB   | Length | Indels | Gaps | Matches               |
|-------------|-------|--|--------|--------|------|-----------------------|
| Qy          | 55    | GAAGGGACGTGCTGATGCCAGAAGGTCACTTGTGAGTCAGTGCCAGGAAAC            | 114    |        |      | Best Local Similarity |
| Db          | 771   | GGGAGAACATGTCATCTTCANGAAGGTGAGTGTCAAGTGTGAGTCAGTGTGAGTCAGGAAAC | 830    | 0;     | 1;   | Conservative          |
| Qy          | 115   | GACACGGGAGGTGGCAAAGACCTCTCTGGTGA-----TGAGCCTAAGGGATGGTT        | 167    |        |      | Mismatches            |
| Db          | 831   | GTACGGGAGGAGGTGGCAAAGACCTCTCTGGTGA-----TGAGCCTAAGGGATGGTT      | 890    | 41;    | 1;   | Indels                |
| Qy          | 168   | GTTGAGGCCCTATGGTTGCACACTGGGATCAGACTCTAACCTCTAACCC              | 218    |        |      | Matches               |
| Db          | 891   | GTACATGGCTCTTAACTCACACATGGGATTGACCTCTATCCTACACT                | 941    |        |      |                       |

RESULT: 13  
US-10-151-081-14  
Sequence 14, Application US/10151081  
Publication No. US20030083295A1

GENERAL INFORMATION:  
APPLICANT: Hadlaczky, Gyula  
Szálay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/151,081  
FILING DATE: 16-MAY-2002  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 09/799 462  
FILING DATE: HERWIRTH 05-MAR-2001  
APPLICATION NUMBER: 09/724 693  
FILING DATE: HERWIRTH 28-NOV-2000  
APPLICATION NUMBER: 08/835,682  
FILING DATE: 10-APR-1997

APPLICATION NUMBER: 08/695,191  
FILING DATE: 07-AUG-1996  
APPLICATION NUMBER: 08/682,080  
FILING DATE: 15-JUL-1996  
APPLICATION NUMBER: 08/629,822  
FILING DATE: 10-AFR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24601-402L  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-151-081-14

| Query Match | Score | DB   | Length | Indels | Gaps                         | Matches            |
|-------------|-------|--|--------|--------|------------------------------|--------------------|
| Qy          | 55    | GAAGGGACGTGCTGATGCCAGAAGGTCACTTGTGAGTCAGTGCCAGGAAAC            | 114    |        | Score 81.4;                  | Length 1400;       |
| Db          | 771   | GGGAGAACATGTCATCTTCANGAAGGTGAGTGTCAAGTGTGAGTCAGTGTGAGTCAGGAAAC | 830    | 0;     | Best Local Similarity 71.9%; | Pred. No. 1.7e-19; |
| Qy          | 115   | GACACGGGAGGTGGCAAAGACCTCTCTGGTGA-----TGAGCCTAAGGGATGGTT        | 167    |        | Mismatches 41;               | Indels 7;          |
| Db          | 831   | GTACGGGAGGAGGTGGCAAAGACCTCTCTGGTGA-----TGAGCCTAAGGGATGGTT      | 890    |        | Matches 123;                 | Gaps 1;            |
| Qy          | 168   | GTTGAGGCCCTATGGTTGCACACTGGGATCAGACTCTAACCTCTAACCC              | 218    |        |                              |                    |
| Db          | 891   | GTACATGGCTCTTAACTCACACATGGGATTGACCTCTATCCTACACT                | 941    |        |                              |                    |

RESULT: 14  
US-10-287-313-14  
Sequence 14, Application US/10287313  
Publication No. US20030101480A1

GENERAL INFORMATION:  
APPLICANT: Hadlaczky, Gyula  
Szálay, Aladar

TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/287,313  
FILING DATE: 01-NOV-2002  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 01/030101480A1-2002  
FILING DATE: <Unknown>  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 09/724,726

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835, 682  
FILING DATE: 28-NOV-2000  
FILING DATE: 10-APR-1997

APPLICATION NUMBER: 08/695, 191  
FILING DATE: 07-AUG-1996

APPLICATION NUMBER: 08/682, 080  
FILING DATE: 15-JUL-1996

APPLICATION NUMBER: 08/629, 822  
FILING DATE: 10-APR-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6869-402N

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: <Unknown>  
ORIGINAL SOURCE:  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-287-313-14

Query Match 35.2% Score 81.4; DB 14; Length 1400;  
Best Local Similarity 71.9%; Pred. No. 1.7e-10;  
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

Qy 55 GGAGGGCAGCTGATTCATGAAGGTCAAGTGTCCCTTAGTTCCCTCCCGGAAAC 114  
Db 771 GGGAGACATGTCATCTTCAGAAGGGTGAAGTGTCCAGCTTCAGGCCAAC 830

Qy 115 GACAGGGAGCTGGCAAGACCTTGATGGTCAAGGGTCAAGGGTGGTT 167  
Db 831 GACACGGGAGCTGGAGTGGTGGCTCAGGGTAAGCCTGTAAGCTTAATCT 890

Qy 168 GTGTAGGGCCCTPATGGTGCACACTGGGATCAAGCTCTACCTCACCC 218  
Db 891 GTACATGGCTCCPTTACACACTGGGATTGACCTCATCTCCACTC 941

Qy 168 GTCAGGGCCCTPATGGTGCACACTGGGATCAAGCTCTACCTCACCC 218  
Db 891 GTACATGGCTCCCTTACCTACACACTGGGATTGACCTCATCTCCACTC 941

RESULT 15  
US-10-219-694-14  
Sequence 14, Application US/10219694  
Publication No. US20030108914A1

GENERAL INFORMATION:  
APPLICANT: Hadlaczky, Gyula  
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS  
NUMBER OF SEQUENCES: 34,  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller, Ehrman White & McAuliffe LLP  
STREET: 4350 La Jolla Village Drive, 7th Floor  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/219, 694  
FILING DATE: 14-Aug-2002  
CLASSIFICATION: <Unknown>

Search completed: October 9, 2003, 23:40:57  
Job time : 208 secs

Gencore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:06:28 : Search time 73 Seconds  
(without alignments)  
1396.706 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtgggtgcaggctaa...ttcacccataggcttgctt 231

Scoring table: IDENTITY\_NUC  
GapOp 10.0 , GapExt 1.0

Searched: 569978 seqs, 20691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:  
1: /cgns2\_6/ptodata/1/ina/5A\_COMB.seq;\*  
2: /cn2\_6/ptodata/1/ina/5B\_COMB.seq;\*  
3: /cgns2\_6/ptodata/1/ina/6A\_COMB.seq;\*  
4: /cgns2\_6/ptodata/1/ina/6B\_COMB.seq;\*  
5: /cgns2\_6/ptodata/1/ina/PCUTS\_COMB.seq;\*  
6: /cgns2\_6/ptodata/1/ina/backfiles1.seq;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query | Match | Length | DB               | ID                 | Description       |                    |
|------------|-------|-------|-------|--------|------------------|--------------------|-------------------|--------------------|
| 1          | 86.2  | 37.3  | 1434  | 3      | US-08-695-191-13 |                    | Sequence 13, Appl |                    |
| 2          | 86.2  | 37.3  | 1434  | 3      | US-08-682-080-13 |                    | Sequence 13, Appl |                    |
| 3          | 81.4  | 35.2  | 1400  | 3      | US-08-695.1.1-14 |                    | Sequence 14, Appl |                    |
| 4          | 81.4  | 35.2  | 1400  | 3      | US-08-682-080-14 |                    | Sequence 14, Appl |                    |
| 5          | 35    | 15.2  | 50    | 4      | US-08-585-593A-4 |                    | Sequence 4, Appl  |                    |
| 6          | 29.2  | 12.6  | 540   | 4      | US-08-250-609-95 |                    | Sequence 95, Appl |                    |
| 7          | 29.2  | 12.6  | 540   | 4      | US-09-250-611-95 |                    | Sequence 95, Appl |                    |
| c          | 8     | 28.6  | 12.4  | 2692   | 1                | US-07-932-454A-2   |                   | Sequence 2, Appl   |
| c          | 9     | 28    | 12.1  | 866    | 3                | US-09-620-12D-682  |                   | Sequence 682, Appl |
| c          | 10    | 27.6  | 11.9  | 1665   | 3                | US-08-766-354A-29  |                   | Sequence 29, Appl  |
| c          | 11    | 27.2  | 11.8  | 2297   | 2                | US-08-588-983-21   |                   | Sequence 21, Appl  |
| c          | 12    | 27.2  | 11.8  | 2297   | 2                | US-08-588-983-21   |                   | Sequence 21, Appl  |
| c          | 13    | 27    | 11.7  | 1054   | 6                | 5189147-4          |                   | Patent No. 5189147 |
| c          | 14    | 26.8  | 11.6  | 503    | 4                | US-09-250-609-85   |                   | Sequence 85, Appl  |
| c          | 15    | 26.8  | 11.6  | 503    | 4                | US-09-250-611-85   |                   | Sequence 85, Appl  |
| c          | 16    | 26.8  | 11.6  | 3213   | 2                | US-08-633-770A-4   |                   | Sequence 4, Appl   |
| c          | 17    | 26.4  | 11.4  | 13865  | 3                | US-09-009-217-11   |                   | Sequence 11, Appl  |
| c          | 18    | 26.4  | 11.4  | 13865  | 3                | US-09-009-656-11   |                   | Sequence 11, Appl  |
| c          | 19    | 26.2  | 11.3  | 71     | 1                | US-08-434-001-210  |                   | Sequence 210, Appl |
| c          | 20    | 26.2  | 11.3  | 71     | 1                | US-08-433-585-210  |                   | Sequence 210, Appl |
| c          | 21    | 26.2  | 11.3  | 71     | 1                | US-08-434-215-210  |                   | Sequence 210, Appl |
| c          | 22    | 26.2  | 11.3  | 71     | 2                | US-08-437-667-210  |                   | Sequence 210, Appl |
| c          | 23    | 26.2  | 11.3  | 71     | 3                | US-08-906-955-210  |                   | Sequence 210, Appl |
| c          | 24    | 26.2  | 11.3  | 71     | 3                | US-08-945-909-210  |                   | Sequence 210, Appl |
| c          | 25    | 26.2  | 11.3  | 71     | 4                | US-09-396-002A-210 |                   | Sequence 210, Appl |
| c          | 26    | 26.2  | 11.3  | 71     | 5                | PCT-US96-06060-210 |                   | Sequence 210, Appl |
| c          | 27    | 26    | 11.3  | 66804  | 4                | US-09-740-041-3    |                   | Sequence 3, Appl   |

## ALIGNMENTS

RESULT 1  
US-08-695-191-13  
; Sequence 13, Application US/08695191  
; Patent No. 602555  
; GENERAL INFORMATION:  
; APPLICANT: Hadlaczky, Gyula  
; ADDRESS: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/695,191  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; APPLICATION NUMBER:  
; FILING DATE: 15-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/629,822  
; FILING DATE: 10-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6869-402C  
; TELECOMMUNICATION INFORMATION:  
; LENGTH: 1434 base pairs  
; TYPE: nucleic acid  
; STRANDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO

; ANTI-SENSE: NO  
 ; FRAGMENT TYPE:  
 ; ORIGINAL SOURCE:  
 US-08-695-191-13

Query Match 37.3%; Score 86.2; DB 3; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 1.3e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

Query Match 37.3%; Score 86.2; DB 3; Length 1434;  
 Best Local Similarity 73.7%; Pred. No. 1.3e-21;  
 Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

Query 55 GGAAAGCACGTCGATGCGATGAGGTCACTGGTCCAGTGTCCAGTCAGGTTCCAGGAAAC 114  
 Db 900 GGGGGACATGTCATCTCATGAGGTCACTGGTCCAGGAAAC 959

Query 55 GGAAAGCACGTCGATGCGATGAGGTCACTGGTCCAGTGTCCAGGAAAC 114  
 Db 900 GGGGGACATGTCATCTCATGAGGTCACTGGTCCAGGAAAC 959

Query 115 GACACGGGGCTGCCAACCTCTCTGGTGA-----TGAGGCTTAAGGGATGGTTT 167  
 Db 960 GACACGGGGCTGCCAACCTCTCTGGTGA-----TGAGGCTTAAGGGATGGTTT 167

Query 115 GACACGGGGCTGCCAACCTCTCTGGTGA-----TGAGGCTTAAGGGATGGTTT 167  
 Db 960 GACACGGGGCTGCCAACCTCTCTGGTGA-----TGAGGCTTAAGGGATGGTTT 167

Query 168 GTGTAGGGCCCTATGCTGCACACTGGGATCAGACTTACCTTACACGGGATTGACCTCTAATCCT 218  
 Db 1020 GTACATGGCTCCCTTAACCTACACGGGATTGACCTCTAATCCT 1070

RESULT 3  
 US-08-695-191-14  
 Sequence 14, Application US/08695191

; GENERAL INFORMATION:  
 ; APPLICANT: Hadlaczky, Gyula  
 ; PATENT NO. 6025155  
 ; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Brown, Martin, Haller & McClain  
 ; STREET: 1660 Union Street  
 ; CITY: San Diego  
 ; STATE: CA  
 ; ZIP: 92101-2926  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FASTSEQ Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/082,080  
 ; FILING DATE: 10-APR-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/629,822  
 ; FILING DATE: 10-APR-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seidman, Stephanie L  
 ; REGISTRATION NUMBER: 33,779  
 ; REFERENCE DOCKET NUMBER: 6869-402B  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-238-0999  
 ; TELEX: 619-238-0062  
 ; INFORMATION FOR SEQ ID NO: 13:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1434 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: Linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; HYPOTHETICAL: NO  
 ; ANTI -SENSE: NO  
 ; FRAGMENT TYPE:  
 ; ORIGINAL SOURCE:  
 US-08-695-191-14

Best Local Similarity 71.9%; Pred. No. 7.3e-20;  
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

Db 771 GGAGGACATGTCACTTCAGAAAGCTTGAGTGTCCAAAGTGTCTCCAGGCCAAC 830  
Qy 115 GAGACGGAGCTGCCAAGACCTCTGGGTA-----TGAGCCTAAGGGATGGTTT 167  
55 GGAGGACATGTCACTTCAGGTCAGTCCCTCCCGGAAAAC 114  
Db 831 GAGACGGAGCTGCCAAGACCTCTGGGTAAGCTGTGAGCTAAGCTTAAGCTCCT 890  
771 GGAGGACATGTCACTTCAGGTCAGTCCCTCCCGGAAAAC 830  
Db 115 GACACGGAGCTGCCAAGACCTCTGGGTA-----TGAGCCTAAGGGATGGTTT 167  
Qy 168 GAGTAGGGCCCTATGCTTGCACACTGGGATCACCTTCACCTTCACCC 218  
Db 891 GAGACATGTCACTTCAGGATGGATTGAGCTTACACACTGGGATTGAGCTCATCTCACTC 941  
831 GAGACGGAGCTGCCAAGACCTCTGGGTAAGCTTAAGCTCCT 890

RESULT 5  
US-08-585-593A-4  
; Sequence 4, Application US/08585593A  
; Patent No. 6503706

GENERAL INFORMATION:

; APPLICANT: ABKEN, Hinrich J  
; APPLICANT: ALBERT, Winfried  
; APPLICANT: JUNGFER, Herbert  
; TITLE OF INVENTION: METHOD OF IDENTIFYING HUMAN AND ANIMAL  
; TITLE OF INVENTION: CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR  
; TITLE OF INVENTION: FORMATION  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-5701

COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible  
; MEDIUM TYPE: FLOPPY disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; COMPUTER: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/585,593A  
; FILING DATE: 16-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP94/02307  
; FILING DATE: 13-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 43 23 727.4  
; FILING DATE: 15-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murray, Robert B  
; REGISTRATION NUMBER: 22,980  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEFAX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 4:  
; LENGTH: 50 base pairs  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; US-08-585-593A-4

Query Match 35.2%; Score 81.4%; DB 4; Length 50;  
Best Local Similarity 100.0%; Pred. No. 0 0015;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 GATCAGACCTCTACCCATGAGGCTTGCCT 231  
Db 1 GATCAGACCTCTACCCATGAGGCTTGCCT 35

RESULT 6  
US-09-250-609-95  
; Sequence 95, Application US/09250609A  
; Patent No. 6458943

Query Match 35.2%; Score 81.4%; DB 3; Length 1400;  
Best Local Similarity 71.9%; Pred. No. 7.3e-20;  
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

Qy 55 GGAAGGGACGCTGTGATTGATGATGAAAGTTCAAGTTCTAGTGTCTAGTCCCAGGAAAAC 114  
Db 1 GATCAGACCTCTACCCATGAGGCTTGCCT 35

Query Match 35.2%; Score 81.4%; DB 3; Length 1400;  
Best Local Similarity 71.9%; Pred. No. 7.3e-20;  
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

Qy 55 GGAAGGGACGCTGTGATTGATGATGAAAGTTCAAGTTCTAGTGTCTAGTCCCAGGAAAAC 114  
Db 1 GATCAGACCTCTACCCATGAGGCTTGCCT 35

GENERAL INFORMATION:  
 APPLICANT: Byrne, Jennifer A.  
 TITLE OF INVENTION: Members of the D52 Gene Family  
 FILE REFERENCE: 1383.0210002  
 CURRENT APPLICATION NUMBER: US/09/250,609A  
 CURRENT FILING DATE: 1999-02-17  
 NUMBER OF SEQ ID NOS: 108  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 95  
 LENGTH: 540  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-250-609-95

**RESULT 7**  
 US-09-250-611-95  
 Sequence 95, Application US/09250611  
 Patent No. 6528283  
 GENERAL INFORMATION:  
 APPLICANT: Byrne, Jennifer A.  
 APPLICANT: Bassett, Paul  
 TITLE OF INVENTION: Members of the D52 Gene Family  
 FILE REFERENCE: 1383.0210001  
 CURRENT APPLICATION NUMBER: US/09/250,611  
 CURRENT FILING DATE: 1999-02-17  
 NUMBER OF SEQ ID NOS: 108  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 95  
 LENGTH: 540  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-250-611-95

Query Match 12.6%; Score 29.2; DB 4; Length 540;  
 Best Local Similarity 54.7%; Pred. No. 0.54;  
 Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 26 CAGAGGATAGCTTGCCTGGCATCCGTGGAAAGGCCACGCTGATGCAAGGTTTCAG 85  
 Db 227 CAGRACCTGTCTAGGACTGTCATGATGAGGTCTACTGCCTACAGAAAGCTCAA 286

Qy 86 TGTCTTAGTTCCTTCCCAGGAAAAACGACACGGGAGCTGGCA 131  
 Db 287 GAAACTCTTACAGGCTGGACAGAAACATCAGCTGCCCTGTCCA 332

**RESULT 7**  
 US-09-250-611-95  
 Sequence 95, Application US/09250611  
 Patent No. 6528283  
 GENERAL INFORMATION:  
 APPLICANT: Byrne, Jennifer A.  
 APPLICANT: Bassett, Paul  
 TITLE OF INVENTION: Members of the D52 Gene Family  
 FILE REFERENCE: 1383.0210001  
 CURRENT APPLICATION NUMBER: US/09/250,611  
 CURRENT FILING DATE: 1999-02-17  
 NUMBER OF SEQ ID NOS: 108  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 95  
 LENGTH: 540  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-250-611-95

Query Match 12.6%; Score 29.2; DB 4; Length 540;  
 Best Local Similarity 54.7%; Pred. No. 0.54;  
 Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 26 CAGAGGATAGCTTGCCTGGCATCCGTGGAAAGGCCACGCTGATGCAAGGTTTCAG 85  
 Db 227 CAGAACCTGTCTAGGACTGTCATGATGAGGTCTACTGCCTACAGAAAGCTCAA 286

Qy 86 TGTCTTAGTTCCTTCCCAGGAAAAACGACACGGGAGCTGGCA 131  
 Db 287 GAAACTCTTACAGGCTGGACAGAAACATCAGCTGCCCTGTCCA 332

**RESULT 8**  
 US-07-932-454A-2/C  
 Sequence 2, Application US/07932454A  
 Patient No. 5262318  
 GENERAL INFORMATION:  
 APPLICANT: GUTHRIE, ELLEN P.  
 TITLE OF INVENTION: ISOLATED DNA ENCODING THE SPH  
 TITLE OF INVENTION: RESTRICTION ENDONUCLEASE AND RELATED METHODS FOR PRODUCING  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
 STREET: 130 WATER STREET  
 ADDRESSEE: CUSHMAN  
 STREET: 130 WATER STREET

CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: US  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/932,454A  
 FILING DATE: 199920820  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WILLIAMS, GREGORY D.  
 REGISTRATION NUMBER: 30901  
 REFERENCE/DOCKET NUMBER: 42078  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 523-3400  
 TELEFAX: (617) 553-6440  
 TELEX: 200291 STRE UR  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2692 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 703..1653  
 OTHER INFORMATION: /note= "METHYLASE GENE STARTS AT POSITION 703/ENDS AT 1653. RESTRICTION ENZYME STARTS AT POSITION 1703/ENDS AT 2410"  
 OTHER INFORMATION:  
 OTHER INFORMATION:  
 OTHER INFORMATION:  
 OTHER INFORMATION:  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1703..2410  
 US-07-932-454A-2

Query Match 12.4%; Score 28.6; DB 1; Length 2692;  
 Best Local Similarity 55.6%; Pred. No. 1.8;  
 Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0

Qy 53 GTGGAAGGCCACGTCTGATTGCCATGAAAGTTCACTGTCTAGTTCCCTTCCCCAGGAAA 112  
 Db 418 GGGGAGTTCCACCGATTCATGCAGTATGTATGCCATGTCATGCCCTTCCTCAAGTCAAG 359

Qy 113 ACGACACGGGAGCTGGCAAGACCTCTCTGGTGAG 151  
 Db 358 GCGGTACGGCACGACGACGACCCCCAACGGGAGCGCTGG 320

**RESULT 9**  
 US-09-620-312D-682/C  
 Sequence 682, Application US/09620312D  
 Patent No. 6569662  
 GENERAL INFORMATION:  
 APPLICANT: Tang, Y. Tom  
 APPLICANT: Liu, Chenghua  
 APPLICANT: Asundi, Vinod  
 APPLICANT: Zhang, Jie  
 APPLICANT: Ren, Feiyan  
 APPLICANT: Chen, Rui-hong  
 APPLICANT: Zhao, Qing A.  
 APPLICANT: Wehrman, Tom  
 APPLICANT: Xue, Aidong J.  
 APPLICANT: Yang, Yonghong  
 APPLICANT: Wang, Jian-Rui  
 APPLICANT: Zhou, Ping  
 APPLICANT: Ma, Yunqing  
 APPLICANT: Wang, Zhiwei  
 APPLICANT: John Fillingham  
 APPLICANT: Drmanac, Radivoje T.  
 APPLICANT: John Fillingham  
 APPLICANT: Wang, Zhiwei  
 APPLICANT: Drmanac, Radivoje T.

; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CTP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIORITY NUMBER: US/09/620,312D  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIORITY OF SEQ ID NOS: 1105  
; SOFTWARE: pl\_FL\_genes Version 1.0  
; SEQ ID NO: 682  
; LENGTH: 866  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (176)..(706)  
; US-09-620,312D-682

Query Match 12.1%; Score 28; DB 4; Length 866;  
Best Local Similarity 51.6%; Pred. No. 1.8;  
Matches 64; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

RESULT 11  
US-08-588-983-21/C  
; Sequence 21, Application US/08588983  
; Patent No. 584067

; GENERAL INFORMATION:  
; APPLICANT: Christopher B. Newgard, et al.  
; TITLE OF INVENTION: Methods and Compositions  
; for Inhibiting Hexokinase  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: US  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/588 983  
; FILING DATE: Concurrently herewith  
; CLASIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fussey, Shelley P.M.  
; REGISTRATION NUMBER: 39,458  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-7300  
; TELEX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2297 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-588-983-21

RESULT 10  
US-08-766-354A-29  
; Sequence 29, Application US/08766354A  
; Patent No. 6013487

; GENERAL INFORMATION:  
; APPLICANT: MITCHELL, LLOYD G.  
; TITLE OF INVENTION: THERAPEUTIC MOLECULES GENERATED BY  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: NIKALDO, MARMELSTEIN, MURRAY, AND ORAM  
; STREET: 655 15TH STREET, N.W., G STREET LOBBY, SUITE  
; STREET: 330  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/766,354A  
; FILING DATE: 13-DEC-1996  
; CLASIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KLESNER, SHARON  
; REGISTRATION NUMBER: 36,335  
; REFERENCE/DOCKET NUMBER: P8159-6-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1665 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; US-08-766-354A-29

Query Match 11.9%; Score 27.6; DB 3; Length 1665;  
Best Local Similarity 48.7%; Pred. No. 3.3;  
Matches 75; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
Qy 66 CTGATTGATGAAGTTCACTGTCCTAGTTCCTAGTCCCTCCAGGAAAGACAGGGAGC 125  
Db 458 CTCACTCCAGCATCTACACCTCTGGCTGCTGCGCCCAACCCGGGTATG 517  
Qy 126 TGGCCAAGACCTCTGGGTATGAGCCPAAGGGATGGTTTGTTGTTAGGGCCCTATGCT 185  
Db 518 AAGCCTGGTACACGGAGGACCAAGGAGATGGATGTTCCAGGTGAAGCTGA 577

Query Match 11.9%; Score 27.6; DB 2; Length 2297;  
Best Local Similarity 49.0%; Pred. No. 5.4;  
Matches 102; Conservative 0; Mismatches 103; Indels 3; Gaps 1;  
Qy 23 GCACRAGGATAAGCTGCTGGTGGAGGCACGCTGATGGCATGAAGGT 82  
Db 2256 GTAGAGGGCTGGATTTGGAAAGGGAGGACTGGGGCTCTGGAGGA 297

Qy 83 CAGTGTCCCTAGTTCCTCCAGGAAAACGACACGGAGCTGGCCNAAGAACCTCTCTG 142

Db 21.96 CACTGAGGGTGGCTTCCTCCCTGGTGAGCTGAGTGTGGAGATTCCTGGCT 2137  
 Qy 143 GGTGATGAGCTTAAGGGATGATGTTGTTGAGGGCCTAT--GCTTGCACAGGGAT 199  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; APPLICATION NUMBER: 620, 122  
 ; FILING DATE: 13-JUN-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 1054  
 5189147-4

RESULT 12  
 US-08-588-976-21/C  
 ; Sequence 21, Application US/085888976  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Christopher B. Newgard, et al.  
 ; TITLE OF INVENTION: Methods and Compositions for  
 ; Inhibiting Hexokinase  
 ; NUMBER OF SEQUENCES: 43  
 ; CURRENT APPLICATION ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: US  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/588, 976  
 ; FILING DATE: Concurrently herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fussey, Shelley P.M.  
 ; REFERENCE/DOCKET NUMBER: UTSD:481/FUS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; TELEX: n/a  
 ; INFORMATION FOR SEQ ID NO: 21:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2297 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-588-976-21

Query Match 11. 8%; Score 27.2; DB 2; Length 2297;  
 Best Local Similarity 49.0%; Pred. No. 5.4; Indels 3; Gaps 1;  
 Matches 102; Conservative 0; Mismatches 103; Indels 3; Gaps 1;  
 Qy 23 GCACAGGGATAGCTGCTGTTGGATCCGTGGAGGCACTGATAGCTGATAGGGT 82  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; APPLICATION NUMBER: 620, 122  
 ; FILING DATE: 13-JUN-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 1054  
 5189147-4

Query Match 11. 7%; Score 27; DB 6; Length 1054;  
 Best Local Similarity 53.3%; Pred. No. 4.5; Indels 0; Gaps 0;  
 Matches 57; Conservative 0; Mismatches 50; Indels 0;  
 Qy 47 CATCCCTGTTGGAAAGGCCAGCTGTGATGATGAAGGGTCAAGTGGCTTCCTGCCCA 106  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 1054  
 5189147-4

Query Match 11. 7%; Score 27; DB 6; Length 1054;  
 Best Local Similarity 53.3%; Pred. No. 4.5; Indels 0; Gaps 0;  
 Matches 57; Conservative 0; Mismatches 50; Indels 0;  
 Qy 803 CCTACTGGGAGGGCCACCCCTATAGCTGTTGGCTGAGCTGGTGTAGTGGTATGGCCAT 862  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 1054  
 5189147-4

Query Match 11. 7%; Score 27; DB 6; Length 1054;  
 Best Local Similarity 53.3%; Pred. No. 4.5; Indels 0; Gaps 0;  
 Matches 57; Conservative 0; Mismatches 50; Indels 0;  
 Qy 107 GGAAAAAACGACGGAGCTGGCAAGACCCCTCTGGTGTAGTGGCC 153  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 1054  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 26 CAGAGGATACTTGTGCTGTTGGCATCTCTGTGAGGGTCTGATTGATGAGTTTCAG 85  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 376 CAGAACCTGTCGTTGGAGCTGGCATGATGTCAGTGGCTCATCTACTGCCTACAGAAGCTCAA 435  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 86 TGTCCTAGTGGCTTCCCTGGAGAAACGACGGAGCTGCGCAAGACCCCTCTG 127  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 436 GAAACTCTTCACAGGTGGACAGAAACATGAGTCGGCTCTG 477  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 143 GGTGATGAGCTTAAGGGATGGTTTGTAGGGCCCTAT--GCTTGCACACTGGGGAT 199  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 2136 TGAATAGTGGAGTCATCTGGCCCTTAGAGCAAGGACTCTGGGCT 2077  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 200 CAGACCTCTACTTCACCCATGAGGCT 227  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

Query Match 11. 6%; Score 26.8; DB 4; Length 503;  
 Best Local Similarity 53.9%; Pred. No. 3.9; Indels 0; Gaps 0;  
 Matches 55; Conservative 0; Mismatches 47; Indels 0;  
 Qy 2076 CTAACCTCTGGCATCTCCTGGAGGCT 2049  
 ; Patent No. 5189147-4  
 ; APPLICANT: SATO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;  
 ; TONIGAWA, SUSUMU  
 ; TITLE OF INVENTION: METERODIMERIC T LYMPHOCYTE RECEPTOR  
 ; ANTIBODY  
 ; NUMBER OF SEQUENCES: 21  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/271, 216  
 ; FILING DATE: 14-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 666, 988  
 ; FILING DATE: 31-OCT-1984  
 ; SEQ ID NO: 4;  
 ; LENGTH: 503  
 5189147-4

; FILE REFERENCE: 1383.0210001  
; CURRENT APPLICATION NUMBER: US/09/250,611  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 85  
; LENGTH: 503  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-250-611-85

Query Match 26 11.68; Score 26.8; DB 4; Length 503;  
Best Local Similarity 53.9%; Pred. No. 3.9;  
Matches 55; Conservative 0; Mismatches 47; Indels 0; Caps 0;

Qy 26 CAGAGGATAGCTGCTGTGGATCCCTGGAGGGCAGTGATGAGGGTCAG 85  
Db 376 CAGAACCTGTCAGGAGCTGGCATGANTGCAGTCTACTGCCPACAAGAGCTCAA 435

Qy 86 TGTCCCTAGTTCCCTTCCCCAGGAAAAACGACACGGAGCTG 127  
Db 436 GAACTCTTCAAGGGTGGACAGAANACATAGCTGCCCTG 477

Search completed: October 9, 2003, 22:44:00  
Job time : 81 secs

**THIS PAGE BLANK (USPCL)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:19:09 ; Search time 3003 Seconds  
(without alignments)  
2551.060 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtgggtgcaggctaa...ttcacccataggcttgctt 231

Scoring table: IDENTRY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 33363688 seqs, 16581889874 residues

Total number of hits satisfying chosen parameters: 66727376

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:  
1: /cgn2\_6/ptodata/1/pna/PCUTUS\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/pna/PCUTUS\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/pna/US0610\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/pna/US0705\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/pna/US080 COMB.seq:\*

6: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq:\*

7: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq:\*

8: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq:\*

9: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq:\*

10: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*

11: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*

12: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq:\*

13: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq:\*

14: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq:\*

15: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq:\*

16: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq:\*

17: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq:\*

18: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq:\*

19: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq:\*

20: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq:\*

21: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq:\*

22: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*

23: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*

24: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*

25: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*

26: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*

27: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq:\*

28: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq:\*

29: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq:\*

30: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq:\*

31: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq:\*

32: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq:\*

33: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq:\*

34: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq:\*

35: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq:\*

36: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq:\*

37: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq:\*

38: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq:\*

39: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq:\*

40: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq:\*

41: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq:\*

42: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq:\*

43: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq:\*

```

44: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US103A_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US103B_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US104A_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US104B_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6023A_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6023B_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
86: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
87: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
88: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
89: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
90: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
91: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
92: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
93: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
94: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
95: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
96: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*
97: /cgn2_6/ptodata/1/pna/US6042_COMB.seq:*
98: /cgn2_6/ptodata/1/pna/US6043_COMB.seq:*
99: /cgn2_6/ptodata/1/pna/US6044_COMB.seq:*
100: /cgn2_6/ptodata/1/pna/US0045_COMB.seq:*
101: /cgn2_6/ptodata/1/pna/US046_COMB.seq:*
102: /cgn2_6/ptodata/1/pna/US047_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query | Match | Length | DB ID               | Description       |
|------------|-------|-------|-------|--------|---------------------|-------------------|
| 1          | 231   | 100.0 | 231   | 29     | US-09-673-716-1     | Sequence 1, Appl1 |
| 2          | 231   | 100.0 | 391   | 29     | US-09-673-716-4     | Sequence 4, Appl1 |
| 3          | 231   | 100.0 | 2052  | 46     | US-10-147-771-25944 | Sequence 25944, A |
| 4          | 231   | 100.0 | 2052  | 91     | US-09-673-716-1     | Sequence 25944, A |

5 231 100.0 2098 46 US-10-144-771-25350 Sequence 25350, A  
 6 231 100.0 2098 91 US-0-360-207-25350 Sequence 25350, A  
 7 231 100.0 2258 46 US-10-144-771-33482 Sequence 33482, A  
 8 231 100.0 2258 91 US-0-360-207-33482 Sequence 33482, A  
 9 231 100.0 2353 46 US-10-144-771-22182 Sequence 22182, A  
 10 231 100.0 2353 91 US-0-360-207-22182 Sequence 22182, A  
 11 231 100.0 3403 46 US-10-144-771-25153 Sequence 25153, A  
 12 231 100.0 3403 91 US-0-360-207-25153 Sequence 25153, A  
 13 231 100.0 5633 46 US-10-144-771-28382 Sequence 28382, A  
 14 231 100.0 5633 91 US-0-360-207-28382 Sequence 28382, A  
 15 231 100.0 5675 46 US-10-144-771-28138 Sequence 28138, A  
 16 231 100.0 62231 45 US-10-144-771-25153 Sequence 493, APP  
 c 18 229.4 99.3 1714 46 US-10-144-771-31295 Sequence 31295, A  
 19 229.4 99.3 1714 91 US-0-360-207-31295 Sequence 31295, A  
 20 229.4 99.3 1891 46 US-10-144-771-35236 Sequence 35236, A  
 21 229.4 99.3 1891 91 US-0-360-207-35236 Sequence 35236, A  
 22 229.4 99.3 1948 46 US-10-144-771-37053 Sequence 37053, A  
 23 229.4 99.3 1948 91 US-0-360-207-37053 Sequence 37053, A  
 24 229.4 99.3 1988 46 US-10-144-771-35849 Sequence 35849, A  
 25 229.4 99.3 1988 91 US-0-360-207-35849 Sequence 35849, A  
 26 229.4 99.3 2049 46 US-10-144-771-29481 Sequence 29481, A  
 27 229.4 99.3 2049 91 US-0-360-207-29481 Sequence 29481, A  
 28 229.4 99.3 2101 46 US-10-144-771-29972 Sequence 29972, A  
 29 229.4 99.3 2246 45 US-10-144-771-25574 Sequence 25574, A  
 30 229.4 99.3 2246 91 US-0-360-207-25574 Sequence 25574, A  
 31 229.4 99.3 2246 99.3 2270 46 US-10-144-771-36364 Sequence 36364, A  
 32 229.4 99.3 2270 91 US-0-360-207-36364 Sequence 36364, A  
 33 229.4 99.3 2472 46 US-10-144-771-28128 Sequence 28128, A  
 34 229.4 99.3 2472 91 US-0-360-207-28128 Sequence 28128, A  
 35 229.4 99.3 2521 46 US-10-144-771-36297 Sequence 36297, A  
 36 229.4 99.3 2521 91 US-0-360-207-36297 Sequence 36297, A  
 37 229.4 99.3 3234 46 US-10-144-771-36142 Sequence 36142, A  
 38 229.4 99.3 3234 91 US-0-360-207-36142 Sequence 36142, A  
 39 229.4 99.3 3234 99.3 17014 46 US-10-144-771-30161 Sequence 30161, A  
 40 229.4 99.3 17014 91 US-0-360-207-30161 Sequence 30161, A  
 41 229.4 98.9 4822 46 US-10-144-771-41694 Sequence 41694, A  
 42 228.4 98.9 4822 91 US-0-360-207-41694 Sequence 41694, A  
 43 228.4 98.9 2084 46 US-10-144-771-34160 Sequence 34160, A  
 44 227.8 98.6 2084 91 US-0-360-207-34160 Sequence 34160, A  
 45 227.8 98.6 2084 96 US-10-144-771-34160 Sequence 34160, A

## ALIGNMENTS

RESULT 1  
 US-09-673-716-1  
 ; Sequence 1, Application US/09673716  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pavlakis, George N.  
 ; APPLICANT: Nappi, Filomena  
 ; APPLICANT: The Government of the United States  
 ; APPLICANT: as represented by the Secretary of  
 ; THE DEPARTMENT OF HEALTH AND HUMAN SERVICES  
 ; TITLE OF INVENTION: Novel Post-Transcriptional Regulatory Elements and Uses  
 ; FILE REFERENCE: 015380-35100US  
 ; CURRENT APPLICATION NUMBER: US/09/673,716  
 ; CURRENT FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: US 60/086,487  
 ; PRIOR FILING DATE: 1998-05-22  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; LENGTH: 231  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence;"rescuing"  
 ; APPLICANT: The Government of the United States  
 ; APPLICANT: as represented by the Secretary of  
 ; THE DEPARTMENT OF HEALTH AND HUMAN SERVICES  
 ; TITLE OF INVENTION: Novel Post-Transcriptional Regulatory Elements and Uses  
 ; FILE REFERENCE: 015380-35100US  
 ; CURRENT APPLICATION NUMBER: US/09/673,716  
 ; CURRENT FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: US 60/086,487  
 ; PRIOR FILING DATE: 1998-05-18  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; LENGTH: 391  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence;"rescuing"  
 ; APPLICANT: The Government of the United States  
 ; APPLICANT: as represented by the Secretary of  
 ; THE DEPARTMENT OF HEALTH AND HUMAN SERVICES  
 ; TITLE OF INVENTION: Novel Post-Transcriptional Regulatory Elements and Uses  
 ; FILE REFERENCE: 015380-35100US  
 ; CURRENT APPLICATION NUMBER: US/09/673,716  
 ; CURRENT FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: US 60/086,487  
 ; PRIOR FILING DATE: 1998-05-22  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; LENGTH: 231  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:core  
 ; OTHER INFORMATION: post-transcriptional regulatory element (PRE)

**RESULT 3**  
US-10-144-771-25944  
; Sequence 25944, Application US/10144771  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 25944  
; LENGTH: 2052  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-25944

Query Match 100.0%; Score 231; DB 46; Length 2052;  
Best Local Similarity 100.0%; Pred. No. 6e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 60  
Db 1088 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 1147

Qy 61 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAGC 120  
Db 1148 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAG 1207

Qy 121 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 180  
Db 1208 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 1267

Qy 181 ATGCTTGCAACTGGGATCAGACCTCTACCCCTACCCCTACCCCTACCCATGGCTTGCTT 231  
Db 1268 ATGCTTGCAACTGGGATCAGACCTCTACCCCTACCCATGGCTTGCTT 1318

**RESULT 5**  
US-10-144-771-25350  
; Sequence 25350, Application US/10144771  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 25350  
; LENGTH: 2098  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-25350

Query Match 100.0%; Score 231; DB 46; Length 2098;  
Best Local Similarity 100.0%; Pred. No. 6e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 60  
Db 1201 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 1260

Qy 61 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAG 120  
Db 1261 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAG 1320

Qy 121 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 180  
Db 1321 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 1380

Qy 181 ATGCTTGCAACTGGGATCAGACCTCTACCCCTACCCCTACCCATGGCTTGCTT 231  
Db 1381 ATGCTTGCAACTGGGATCAGACCTCTACCCATGGCTTGCTT 1431

**RESULT 6**  
US-60-360-207-25350  
; Sequence 25350, Application US/60360207  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 25350  
; LENGTH: 2098  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-25350

Query Match 100.0%; Score 231; DB 91; Length 2098;  
Best Local Similarity 100.0%; Pred. No. 6e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 60  
Db 1201 GTGGGTGCGAGGCTAACGACTGCACAGAGGATAGCTTGCTGTCCTGGAAAGG 1260

Qy 61 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAG 120  
Db 1261 CACGTCTGATTGCTGATGAGGTTCAAGTGCCTGCTAGTCCTCCAGAAAACGACAG 1320

Qy 121 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 180  
Db 1321 GGAGCTGCCAAGACCTCTGGGTGATGAGCTTAAGGATGGTTTGTAGGGCCCT 1380

Qy 181 ATGCTTGCAACTGGGATCAGACCTCTACCCCTACCCATGGCTTGCTT 231

Db 1381 ATGCTTGACACTGGGATCAGACCTTACCTTCACCATGAGCTTGCCT 1431  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO: 33482  
; LENGTH: 2258  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-33482

Query Match Score 231; DB 46; Length 2258;  
Best Local Similarity 100.0%; Pred. No. 6.2e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Gaps 0;  
Qy 1 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 60  
Db 1299 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 1358  
Qy 61 CACGCTGATGATGATGAAGGTCAGTGTCCTAGTCCCTCCCCAGAAAAAGACAG 120  
Db 1359 CACGCTGATGATGATGAAGGTCAGTGTCCTAGTCCCTCCCCAGAAAAAGACAG 1418  
Qy 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 180  
Db 1419 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 1478  
Qy 181 ATGCTTGACACTGGGATCAGACCTTACCTTACCATGAGCTTGCCT 231  
Db 1479 ATGCTTGACACTGGGATCAGACCTTACCTTACCATGAGCTTGCCT 1529

RESULT 8  
US-60-360-207-33482  
; Sequence 33482, Application US/60360207  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO: 33482  
; LENGTH: 2258  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-33482

Query Match Score 231; DB 91; Length 2258;  
Best Local Similarity 100.0%; Pred. No. 6.2e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Gaps 0;  
Qy 1 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 60  
Db 1299 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 1358  
Qy 61 CACGCTGATGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 120  
Db 1359 CACGCTGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 1418  
Qy 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 180  
Db 1419 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 1478

RESULT 9  
US-10-144-771-22182  
; Sequence 22182, Application US/10144771  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/10/144,771  
; CURRENT FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO: 22182  
; LENGTH: 2353  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-10-144-771-22182

Query Match Score 231; DB 46; Length 2353;  
Best Local Similarity 100.0%; Pred. No. 6.3e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Gaps 0;  
Qy 1 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 60  
Db 1328 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 1387  
Qy 61 CACGCTGATGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 120  
Db 1388 CACGCTGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 1447  
Qy 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 180  
Db 1448 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 1507  
Qy 181 ATGCTTGACACTGGGATCAGACCTTACCTTACCATGAGCTTGCCT 231  
Db 1508 ATGCTTGACACTGGGATCAGACCTTACCTTACCATGAGCTTGCCT 1558

RESULT 10  
US-60-360-207-22182  
; Sequence 22182, Application US/60360207  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO: 22182  
; LENGTH: 2353  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-22182

Query Match Score 231; DB 91; Length 2353;  
Best Local Similarity 100.0%; Pred. No. 6.3e-65; Mismatches 0; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Gaps 0;  
Qy 1 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 60  
Db 1328 GTGGGTGCAAGGTAAGCAGTGCACAGGATAGCTGTGTTGGCAAGG 1387  
Qy 61 CACGCTGATGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 120  
Db 1388 CACGCTGATGATGAAGGTCAGTGTCCCTCCCCAGAAAAAGACAG 1447  
Qy 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 180  
Db 1448 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGTTGGTAGGCCCT 1507

Qy 181 ATGCTTGACACTGGGATCAGACCTTACCCCTAACCGTAGGGCTTGCTT 231  
 Db 1508 ATGCTTGACACTGGGATCAGACCTTACCCCTAACCGTAGGGCTTGCTT 1558  
 Db 1717 ATGCTTGACACTGGGATCAGACCTTACCCCTAACCGTAGGGCTTGCTT 1767

RESULT 11  
 US-10-144-771-25153  
 ; Sequence 25153, Application US/10144771  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig  
 ; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
 ; FILE REFERENCE: CLO01321  
 ; CURRENT APPLICATION NUMBER: US/10/144-771  
 ; CURRENT FILING DATE: 2002-05-15  
 ; NUMBER OF SEQ ID NOS: 47235  
 ; SEQ ID NO: 25153  
 ; LENGTH: 3403  
 ; TYPE: DNA  
 ; ORGANISM: HUMAN  
 ; US-10-144-771-25153

Query Match Score 231; DB 46; Length 3403;  
 Best Local Similarity 100.0%; Pred. No. 7e-65; Mismatches 0;  
 Matches 231; Conservative 0; Indels 0; Gaps 0;

Qy 1 GTGGGCTGGAGGCTTAAGGAACTGCAAGGATTAGCTTGCTGTGGATCCCTGTGGAGG 60  
 Db 1537 GTGGGCTGGAGGCTTAAGCAGCTCACAGAGGATAAGCTGTGTGGATCCCTGTGGAGG 1596  
 Qy 61 CAGCTCTGATTGATGTAAGGTTCTAGTGTGCTAGTCCTTCCCAGAAAAGCACAG 120  
 Db 1597 CAGCTCTGATTGATGTAAGGTTCTAGTGTGCTAGTCCTTCCCAGAAAAGCACAG 1656  
 Qy 121 GGACTTGCCAAAGACCTCTGGGTGATGAGCTTAAGGGATGGTTGTAGGGCCCT 180  
 Db 1657 GGACTTGCCAAAGACCTCTGGGTGATGAGCTTAAGGGATGGTTGTAGGGCCCT 1716  
 Qy 181 ATGCTTGACACTGGGATCAGACCTTACCCCTAACCGTAGGGCTTGCTT 231  
 Db 1717 ATGCTTGACACTGGGATCAGACCTTACCCCTAACCGTAGGGCTTGCTT 1767

RESULT 12  
 US-60-360-207-25153  
 ; Sequence 25153, Application US/60360207  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig  
 ; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
 ; FILE REFERENCE: CLO01321  
 ; CURRENT APPLICATION NUMBER: US/60/360-207  
 ; CURRENT FILING DATE: 2002-03-01  
 ; NUMBER OF SEQ ID NOS: 47235  
 ; SEQ ID NO: 25153  
 ; LENGTH: 3403  
 ; TYPE: DNA  
 ; ORGANISM: HUMAN  
 ; US-60-360-207-25153

Query Match Score 231; DB 91; Length 3403;  
 Best Local Similarity 100.0%; Pred. No. 7e-65; Mismatches 0;  
 Matches 231; Conservative 0; Indels 0; Gaps 0;

Qy 1 GTGGGCTGGAGGCTTAAGGAACTGCAAGGATTAGCTTGCTGTGGATCCCTGTGGAGG 60  
 Db 1537 GTGGGCTGGAGGCTTAAGCAGCTCACAGAGGATAAGCTGTGTGGATCCCTGTGGAGG 1596  
 Qy 61 CAGCTCTGATTGATGTAAGGTTCTAGTGTGCTAGTCCTTCCCAGAAAAGCACAG 120  
 Db 1597 CAGCTCTGATTGATGTAAGGTTCTAGTGTGCTAGTCCTTCCCAGAAAAGCACAG 1656  
 Qy 121 GGACTTGCCAAAGACCTCTGGGTGATGAGCTTAAGGGATGGTTGTAGGGCCCT 180

```

Db      1761 GGAGCTGGCCAAGACCTCTGGGATGAGCCTAAGGGATGGTTAGGCCCT 1820
Qy      181 ATGCCTGACACTGGGATCAGACCTCACCCATGAGGCTTGCCT 231
Db      1821 ATGCCTGACACTGGGATCAGACCTCACCCATGAGGCTTGCCT 1871

RESULT 15
US-10-144-771-28138
; Sequence 28138, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO: 28138
; LENGTH: 5675
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5675)
; OTHER INFORMATION: n = A,T,C or G
US-10-144-771-28138

Query Match          100.0% ; Score 231; DB 46; Length 5675;
Best Local Similarity 100.0% ; Pred. No. 8.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GTGGGTGGAGGCTAAGCACTGGCACAGGATAAGCTGTCATCCTGTGGAAAGG 60
Db      1301 GTGGGTGGAGGCTAAGCACTGGCACAGGATAAGCTGTCATCCTGTGGAAAGG 1360
Qy      61 CACGCTGATGCGAAGGTCACTGTCAGTCCAGAGGATAAGCTGTCATCCTGTGGAAAGG 120
Db      1361 CACGCTGATGCGAAGGTCACTGTCAGTCCAGAGGATAAGCTGTCATCCTGTGGAAAGG 1420
Qy      121 GGAGCTGGCCAAGACCTCTGGGATGAGCCTAAGGGATGGTTAGGCCCT 180
Db      1421 GGAGCTGGCCAAGACCTCTGGGATGAGCCTAAGGGATGGTTAGGCCCT 1480
Qy      181 ATGCCTGACACTGGGATCAGACCTCACCCATGAGGCTTGCCT 231
Db      1481 ATGCCTGACACTGGGATCAGACCTCACCCATGAGGCTTGCCT 1531

```

Search completed: October 9, 2003, 23:34:11  
 Job time : 3004 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:41:03 ; Search time 171 Seconds  
(without alignments)  
1886.736 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gttgggtgcggctaaagca.....ttcacccatgaggcttgctt 231

Scoring table: IDENTITY\_NUC Gapop 10.0 , Gapext 1.0

Searched: 1776535 seqs, 698337343 residues

Total number of hits satisfying chosen parameters: 3553070

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New;\*

1: /cgns2\_6/ptodata/1/pna/us06\_new\_comb.seq;\*

2: /cn2\_6/ptodata/1/pna/us07\_new\_comb.seq;\*

3: /cgns2\_6/ptodata/1/pna/us07\_new\_comb.seq;\*

4: /cgns2\_6/ptodata/1/pna/us08\_new\_comb.seq;\*

5: /cn2\_6/ptodata/1/pna/us09\_new\_comb.seq;\*

6: /cgns2\_6/ptodata/1/pna/us10\_new\_comb.seq;\*

7: /cgns2\_6/ptodata/1/pna/us60\_new\_comb.seq;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Description                                   |   |
|------------|-------|-------------|---|---|
|            |       | %           | DB ID   |   |
| 1          | 72    | 31.2        | 96593 1 PCT-US02-38582-151 Sequence 151, App  |   |
| 2          | 35    | 15.2        | 30593 7 US-60-495-16231 Sequence 16231, A     |   |
| c          | 3     | 14.7        | 49596 7 US-60-495-135-164 Sequence 3644, Ap   |   |
| 4          | 32.6  | 14.1        | 1106 6 US-10-425-114A-16270 Sequence 16220, A |   |
| 5          | 32.6  | 14.1        | 1145 6 US-10-425-114A-15489 Sequence 25489, A |   |
| 6          | 31    | 13.4        | 1205 6 US-10-425-114A-16110 Sequence 16110, A |   |
| 7          | 31    | 13.4        | 1211 6 US-10-425-114A-23937 Sequence 23937, A |   |
| 8          | 30    | 13.1        | 25995 7 US-60-500-337-19598 Sequence 19598, A |   |
| 9          | 29.8  | 12.9        | 37637 7 US-60-500-315-11368 Sequence 11368, A |   |
| 10         | 29.6  | 12.8        | 959 6 US-10-663-561-52 Sequence 452, App      |   |
| 11         | 29.6  | 12.8        | 1531 6 US-10-425-114A-1045 Sequence 1045, Ap  |   |
| 12         | 29.6  | 12.8        | 2731748 6 US-10-297-465B-1 Sequence 1, Appl.  |   |
| c          | 13    | 29.4        | 12.7  | 19946 7 US-60-500-337-19711 Sequence 19711, A |
| 14         | 28.8  | 12.5        | 34227 7 US-60-500-15-11594 Sequence 11594, A  |   |
| c          | 15    | 28.4        | 12.3  | 25821 7 US-60-487-610-2089 Sequence 20089, A  |
| c          | 16    | 28.4        | 12.3  | 25821 7 US-60-485-50-12348 Sequence 12348, A  |
| c          | 17    | 28.2        | 12.2  | 15421 7 US-60-485-50-11760 Sequence 11760, A  |
| c          | 18    | 27.8        | 12.0  | 201 7 US-60-485-50-10495 Sequence 10495, A    |
| 19         | 27.8  | 12.0        | 201 7 US-60-485-450-10505 Sequence 10505, A   |   |
| 20         | 27.8  | 12.0        | 201 7 US-60-485-50-10515 Sequence 10515, A    |   |
| 21         | 27.8  | 12.0        | 201 7 US-60-485-50-11518 Sequence 41518, A    |   |
| 22         | 27.8  | 12.0        | 34570 1 PCT-US02-38582-157 Sequence 157, App  |   |
| 23         | 27.8  | 12.0        | 45945 7 US-60-500-337-19316 Sequence 19316, A |   |
| c          | 24    | 27.8        | 12.0  | 85568 7 US-60-495-114-16789 Sequence 16789, A |
| c          | 25    | 27.8        | 12.0  | 85568 7 US-60-495-135-3707 Sequence 3707, Ap  |
| 26         | 27.8  | 12.0        | 329776 5 US-09-947-914-49 Sequence 49, Appl   |   |

## ALIGNMENTS

RESULT 1  
PCT-US02-38582-151  
; Sequence 151, Application PC/TUS0238582

; GENERAL INFORMATION:  
; APPLICANT: SAGRES DISCOVERY  
; APPLICANT: MORRIS, David W.  
; APPLICANT: ENGELHARD, ERIC K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 52945100143  
; CURRENT APPLICATION NUMBER: PCT/TUS02/38582  
; CURRENT FILING DATE: 2003-06-03  
; PRIORITY APPLICATION NUMBER: 09/937,722  
; PRIORITY FILING DATE: 2001-11-30  
; NUMBER OF SEQ ID NOS: 301  
; SOFTWARE: FastSeqD for Windows Version 4.0  
; SEQ ID NO: 151  
; LENGTH: 90593

; ORGANISM: Mus Musculus  
; TYPE: DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(96593)  
; OTHER INFORMATION: n = A,T,C or G  
PCT-US02-38582-151

Query Match Score 72; DB 1; Length 96593;  
Best Local Similarity 72.1%; Pred. No. 6.7e-14;  
Matches 124; Conservative 0; Mismatches 40; Indels 8; Gaps 2;

Qy 55 GGAAAGCACTGTCATTGGCATGTCATGCTGAGTGTGACTGGCCCAAATCGCTCTCCAGCGCAA 113  
Db 1044 GGAGACATGTCATCTTCAGAAAGTGTGACTGGCCCAAATCGCTCTCCAGCGCAA 1103

Qy 114 CGACAGGGAGCTGGCCAAGACCTCTCTGGGTA-----TGAGCCTAAGGGATGGTT 166  
Db 1104 CGACAGGGAGCTGGCTCTGGGTAAGGCCTGAGCTTAAGGCTAATCC 1163

Qy 167 TGTGAGGGCCCTATGCTGACACTGGGATGCAACCTCTACCTTCACCC 218  
Db 1164 TGTACATGGCTCTTACCTGCACACTGGGATGTCACCTCTACCC 1215

RESULT 2  
US-60-495-114-16231  
; Sequence 16231, Application US/60495114  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
TITLE OF INVENTION: USES THEREOF  
FILE REFERENCE: CLO01480  
CURRENT APPLICATION NUMBER: US/60/495,114  
CURRENT FILING DATE: 2003-08-15  
NUMBER OF SEQ ID NOS: 91238  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO: 16231  
LENGTH: 30593  
TYPE: DNA  
ORGANISM: Homo sapiens  
-60-495-114-6231

| Query   | Match  | Score | DB         | Length |
|---------|--|-------|------------|--------|
| Best    | Local  | 15.2% | 35;        | 30593; |
| Matches | Similarity   | 55.3% | Pred. No.  | 0.21;  |
| 68;     | Conservative   | 0;    | Mismatches | 55;    |
|         |  |       | Indels     | 0      |
| 4       | GGTGTGGCTTAAGCACTGCAAGGATAGCTTGCGATTCCTGTT               |       |            |        |
| 15181   | GGTGGGGGGACAGTCAGGGCCAGGGACTCAGAGCTCTGCTGCATGGG          |       |            |        |
| 64      | GCTGTATTGCTCATGAGGTTCAGTGTGTCCTAGTTCCCCTCCCCAGGGAAAAACCC |       |            |        |
| 15241   | GTTGGGATTTGGGATCTGCCCTGGCAGACCCCAACCCCAAGCCACAGG         |       |            |        |
| 124     | GCT  | 126   |            |        |
| 15301   | GCT  | 15303 |            |        |

SU LT 3  
-60-495-135-3644/c  
Sequence 3644, Application US/60495135  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS,  
METHODS OF  
TITLE OF INVENTION: USES THEREOF  
FILE REFERENCE: CL001479  
CURRENT APPLICATION NUMBER: US/60/495,135  
CURRENT FILING DATE: 2003-08-15  
NUMBER OF SEQ ID NOS: 18339  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO: 3644  
LENGTH: 4,9556  
TYPE: DNA  
ORGANISM: Homo sapiens

| Query              | Match  | Score      | DB  | Length |
|--------------------|--|------------|---|--------|
| Best Local Matches | 61;  | 34;        | 49596;  |        |
| Local Similarity   | 57.5%  | Pred.      | No.   | 0.52;  |
| Conservative       | 0;   | Mismatches |   | 0      |
| Matches            | 45;  | Indels     |   |        |
|                    |  |            |   |        |
| 11                 | AGGCTAACACACTGACAGAGGA                           | 94         | CAGCTGTGGATGCCGAAAGGA                         |        |
| 4553               | AGGCACAGCTGGGACTCACTATGTTGCTACCGGCATATTAGCGTGGGA | 71         | TGCGTGGGTTCACTATGTTGCTAGTGTCCCTCCCGAGAAAAACGA | 116    |
| 4552               | TTGGCTTGCTGGGACTCACTATGTTGCTACCGGCATATTAGCGTGGGA | 71         | TGCGTGGGTTCACTATGTTGCTAGTGTCCCTCCCGAGAAAAACGA | 116    |

SUITE 4  
-10-425-114A-16270  
Sequence 16270 Application #/10A2511A

GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovacic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5331)B  
; CURRENT APPLICATION NUMBER: US/10/4225,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO: 16270  
; LENGTH: 1106  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3062-025-A9\_FI.I  
rc: 10425,114,16270

|         |       |   |        |            |     |        |        |       |
|---------|-------|---|--------|------------|-----|--------|--------|-------|
| Query   | Match | 14.18;  | Score  | 32.6;      | DB  | 6;     | Length | 1106; |
| Best    | Local | Similarity  | 52.68; | Pred.      | No. | 0.51;  |        |       |
| Matches | 71;   | Conservative  | 0;     | Mismatches | 64; | Indels | 0;     | Gaps  |
| Qy      | 31    | GATAGCTTGCCTGGCATCCGTGAAAGGCCACGTCGTATTGCAATGAAGGTTCAAGTGTC | 90     |            |     |        |        |       |
| Db      | 715   | GCTAGATGGTCAGCCAAACGGACCGGAAAGCTGGATGGTTCTGACAAAGCGGGCA     | 77     |            |     |        |        |       |
| Qy      | 91    | TAGTCCCTTCCCCGGAAAACGACACGGGAGCTGGCAAGACCTCTGGGTGATGA       | 15     |            |     |        |        |       |
| Db      | 775   | TCACTCCTGACTCCAGAACGACGACGACGACGACGACGACGACGACGACGACGACG    | 83     |            |     |        |        |       |
| Qy      | 151   | GCCTAAGGGATGGTT   | 165    |            |     |        |        |       |
| Db      | 835   | ACTCTGAGGATGGTT   | 849    |            |     |        |        |       |

RESULT 5  
US-10-425-114A-25489  
; Sequence 25489, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with Plants and Uses Thereof for Plant Improvement  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425, 114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO: 25489

```

; LENGTH: 1145
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-015-C11_FLI
; US-10-425-114A-25469

Query Match Score 32.6; DB 6; Length 1145;
Best Local Similarity 52.6%; Pred. No. 0.52;
Matches 71; Conservative 0; Mismatches 64; Indels 0;
Gaps 0
31 GATACTTGGTGTGGCATCTGTGGAAGGCCACGCTGATTGATGAAAGTTCACTGTGCC
Qy

```

| Db  | Qy   | 91 | TACITCCCTCCCCAGGAAAAACGACAGGGAGCTGGATGGTTCTGGACAGGGGGGCA | 81 |
|-----|--|----|--|----|
| 753 | GCTAATGGTTCGCCAACGGACGGAAAGCTGGATGGTTCTGGACAGGGGGGCA | 15 |  |    |
| 812 | TTCTGGACAGGGGGGCA                                    | 1  |  |    |
| 813 | TTCTGGACAGGGGGC                                      | 1  |  |    |

|    |     |                   |     |
|----|-----|-------------------|-----|
| Qy | 151 | GCCCTAAAGGGATGGTT | 165 |
|    |     |                   |     |
| Db | 873 | ACTCTGAGGATCGTT   | 887 |

RESULT 6  
US-10-425-114A-16110

; Sequence 16110, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Sreen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 3B-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO: 16110  
; LENGTH: 1205  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3061-082-A12\_FLI

Query Match 13.4%; Score 31; DB 6; Length 1205;  
Best Local Similarity 51.9%; Pred. No. 1.8;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 31 GATAGCTTGCTGTTGGCATCCTGTGGAGGCCAGCTGATTCATGAGGTTACAGTGTCC 90  
Db 815 GCTAGATAATGGTCCAGCAACGGACGGTGTGTTCTGGACAAGGGGGCGCA 874

Qy 91 TAGTTCCCTTGGCAGAAACGAAACGGAACGGGAGCTGGTGTGATGTA 150  
Db 875 TCATTCTTGACTCCAAGAACGACGATGACGAGCTTGCTAGAGGCTCATAGTGTCA 934

Qy 151 GCCTAAGGGATGGTT 165  
Db 935 ACTCCGAGGATCCT 949

RESULT 7  
US-10-425-114A-23937  
; Sequence 23937, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Sreen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 3B-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO: 23937  
; LENGTH: 1211  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3606-028-C5\_FLI

Query Match 13.4%; Score 31; DB 6; Length 1211;  
Best Local Similarity 51.9%; Pred. No. 1.8;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 31 GATAGCTTGCTGTTGGCATCCTGTGGAGGCCAGCTGATTCATGAGGTTACAGTGTCC 90  
Db 787 GCTGAGATGGTCCAGAACGACGGAGCTGGTGTAGGTGTGAGTCATATTTCCTTCACANGGA 846

Qy 91 TAGTTCCCTTCCCCAGAAAAACGACACGGAGCTGGCCAGACCTCTCTGGGTGTGA 150

RESULT 8  
US-60-500-337-19598  
; Sequence 19598, Application US/60500337  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND USES THEREOF  
; TITLE OF INVENTION: CURENT APPLICATION NUMBER: US/60/500,337  
; CURRENT FILING DATE: 2003-09-05  
; NUMBER OF SEQ ID NOS: 123188  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 19598  
; LENGTH: 25995  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-60-500-337-19598

Query Match 13.1%; Score 30.2; DB 7; Length 25995;  
Best Local Similarity 53.9%; Pred. No. 8.6;  
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 59 GGAGCTGCTGATTCATGAGGTTACAGTGTCCTAGTTCCCTCCAGGAAAACGCGA 118  
Db 3083 GGAGGGCGAGGGCTAGGGTCACTGGTCACTGGTCACTGGTCACTGGTCACTGGTCA 3142

Qy 119 CGGGAGCTGGCCAAAGACCTCTCTGGGTATGAGCCCTAAGGGATGGTTTGTGTAG 173  
Db 3143 CAGSCGCTGGGACTGGCCGGCTGGTCACTGGGAGCTGGGAGCTGGAGGGAG 3197

RESULT 9  
US-60-500-315-11368  
; Sequence 11368, Application US/60500315  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS OF DETECTION AND USES THEREOF  
; TITLE OF INVENTION: CURENT APPLICATION NUMBER: US/60/500,315  
; CURRENT FILING DATE: 2003-09-05  
; NUMBER OF SEQ ID NOS: 63978  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 11368  
; LENGTH: 37637  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-60-500-315-11368

Query Match 12.9%; Score 29.8; DB 7; Length 37637;  
Best Local Similarity 63.0%; Pred. No. 13;  
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 51 CTGGAAAGGCACGTCATTGATGAAAGTTCACTTCCCTCCAGAA 110  
Db 36010 CTTTGGGAGGAAGTCCTTGTGTCATATTTCCTTCACANGGA 36069

Qy 111 AACGACACGGGA 123  
Db 36070 TCAAGTCACTGTA 36082

RESULT 10  
US-10-663-561-452

Sequence 452, Application US/10663561  
; GENERAL INFORMATION:  
; APPLICANT: DENSLAW, NANCY D.  
; APPLICANT: LARKIN, PATRICK M.  
; APPLICANT: FOLMAR, LEROY C.  
; APPLICANT: HEMMER, MICHAEL J.  
; TITLE OF INVENTION: DETECTING HORMONALLY ACTIVE COMPOUNDS  
; CURRENT APPLICATION NUMBER: US/10/663,561  
; CURRENT FILING DATE: 2003-09-15  
; NUMBER OF SEQ ID NOS: 566  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO: 452  
; LENGTH: 959  
; TYPE: DNA  
; ORGANISM: Micropterus salmoides  
; US-10-663-561-452

Query Match 12.8%; Score 29.6; DB 6; Length 959;  
Best Local Similarity 56.0%; Pred. No. 5.2; Indels 0; Gaps 0;  
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

|    |   |    |  |
|----|---|----|--|
| Qy | 103 CCCAGGAAAAACGACACGGAGCTGGCAAGACCTCTGGGTATGAGCCTAAGGGATG 162 | Db | 420 CTCAGAAAAAAGAGAGATGGGGCTGGACGGAGGGCTGGCCAAAGGTGT 479 |
| Qy | 163 GTTTTGTTAGGGCCCTATGCTTGCACACTGGGATAG 202                    | Db | 480 GGTTCTGTTGGACCTCTGATCATGGCCTTCTGAAACAG 519           |

RESULT 11  
; Sequence 1045, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; Title of Invention: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO: 1045  
; LENGTH: 1531  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700100671\_FLI  
; US-10-425-114A-1045

Query Match 12.8%; Score 29.6; DB 6; Length 1531;  
Best Local Similarity 54.6%; Pred. No. 5.9; Indels 0; Gaps 0;  
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

|    |   |    |  |
|----|---|----|--|
| Qy | 56 GAAGGCACTCTGATTGGCATGAGGTTAGTGTCTCTAGTCTCCCTCCAGGAAACG 115 | Db | 157 GAAGTCCATGCTGACCCATTAAAGTCAGCGTCCGACAAGGTCCCAGAAGGAG 216 |
| Qy | 116 ACAGGGACCTGGCAAAGACCTCTGGTATGAGCTAAGGGATGG 163            | Db | 217 AATGGGAAATGGAACGACGAAGGTTGATGAGCCAAAGGAATTG 264          |

RESULT 12  
; Sequence 11594, Application US/60500315  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS  
; OF DETECTION AND USES THEREOF

RESULT 13  
US-60-500-337-19711/C  
; Sequence 19711, Application US/60500337  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01483  
; CURRENT APPLICATION NUMBER: US/60/500,337  
; CURRENT FILING DATE: 2003-09-05  
; NUMBER OF SEQ ID NOS: 122188  
; SEQ ID NO: 19711  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; LENGTH: 19946  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-60-500-337-19711

Query Match 12.7%; Score 29.4; DB 7; Length 19946;  
Best Local Similarity 54.1%; Pred. No. 15; Indels 0; Gaps 0;  
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

|    |   |    |  |
|----|---|----|--|
| Qy | 67 TGATGGCATGAAAGTTTCACTGTCTAGTCTCCCTCCAGGAAACACCCGAGCT 126 | Db | 662 TCACTGCGGAAGGGGCCCTCGAGGCTCCCTCCAGCCCCACATCTGGGGCT 603 |
| Qy | 127 GGCAAGAACCTCTGGGTATGAGCTTGTAGGGATGTTGTAGGGCC 177        | Db | 602 GGCCCCAGGATAGTGGGAGGAGGCTGTGCTGAGCTCGTAC 552           |

RESULT 14  
; Sequence 11594, Application US/60500315  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS  
; OF DETECTION AND USES THEREOF

```

; FILE REFERENCE: CL001484
; CURRENT APPLICATION NUMBER: US/60/500,315
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 69978
; SOFTWARE: Fastaseq for Windows Version 4.0
; SEQ ID NO: 11594
; LENGTH: 34227
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-11594

Query Match      12.5%;  Score 28.8;  DB 7;  Length 34227;
Best Local Similarity  58.1%;  Pred. No. 28;
Matches 51;  Conservative 0;  Mismatches 37;  Indels 0;  Gaps 0;
Oy          115 GACACGGAGCTGGCCAGAACGCTCTGGGTATGACCTAAGGGAACTGGTGTAGG 174
Db          17633 GAGGCCTGGCTGCGAGGCCAGAGGGAGGCTAAGGAATGGCAGGGCTAAGGAATGGCAGGGCTCAGG 17692
Oy          175 GCCCCTATGCTTCACACTGGGATCA 202
Db          17693 TCCCAAAGCCCTGCCACAGGGAGGAG 17720

```

---

```

RESULT 15
US-60-487-610-20089/c
; Sequence 20089, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001489
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: Fastaseq for Windows Version 4.0
; SEQ ID NO: 20089
; LENGTH: 25821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-20089

Query Match      12.3%;  Score 28.4;  DB 7;  Length 25821;
Best Local Similarity  58.1%;  Pred. No. 35;
Matches 50;  Conservative 0;  Mismatches 36;  Indels 0;  Gaps 0;
Oy          51 CNGTGGAAAGGCAAGTCTGATTCATGAGGGTCAGTTCCCTCCCAAGAA 110
Db          22039 CTAGGGGATGCTGGTCATGGTGAAGGGTAGGGTCGGGAAGCTGGCAGGCC 22780
Oy          111 AACGACACGGAGCTGGCAAGACC 136
Db          22779 GATGCAACCGGACCTGGTCAAGGCC 22754

```

Search completed: October 9, 2003, 23:37:22  
Job time : 179 secs

**THIS PAGE BLANK (USP TO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 20:13:18 ; Search time 256 Seconds  
(without alignments)

2435.821 Million cell updates/sec  
2435.821 Million cell updates/sec

Title: US-09-673-716-1  
Perfect score: 231  
Sequence: 1 gtgggtgcgaggtaagca.....ttcacccataggcttgctt 231

Scoring table: IDENTITY\_NUC

Gapon 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:

5105512

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:  
 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT; \*  
 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT; \*  
 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT; \*  
 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT; \*  
 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT; \*  
 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT; \*  
 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT; \*  
 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT; \*  
 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT; \*  
 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT; \*  
 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT; \*  
 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT; \*  
 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT; \*  
 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT; \*  
 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT; \*  
 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT; \*  
 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT; \*  
 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT; \*  
 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT; \*  
 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT; \*  
 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT; \*  
 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT; \*  
 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT; \*  
 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT; \*  
 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT; \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query | Match | Length | DB ID    | Description  |
|------------|-------|-------|-------|--------|----------|--|
| 1          | 231   | 100.0 | 231   | 21     | AZ255983 | Core fragment B, a Murine genomic DNA sequence of ne             |
| 2          | 231   | 100.0 | 393   | 21     | AZ255986 | Nucleotide sequence of ne  |
| 3          | 224   | 97.0  | 226   | 21     | AZ255987 | Nucleotide sequence of ne  |
| 4          | 114.4 | 49.5  | 75798 | 25     | ABX77212 | DNA probe 3 detect   |
| c          | 89.4  | 38.7  | 18910 | 25     | ABZ80240 | Human histone d2c  |
| 5          | 89.4  | 38.7  | 5018  | 6      | AAM50150 | Human ovarian cancer   |
| 6          | 89.4  | 38.7  | 5018  | 8      | AAN70646 | Human Colon cancer   |
| 7          | 89.4  | 38.7  | 3336  | 6      | AAN50149 | Murine Post translational regulatory element; PRB; murine; NCTE; |
| 8          | 87.8  | 38.0  | 30    | 21     | AAS55984 | Post-transcriptional RNA nucleo-cytoplasmic transport element;   |

| Result No. | Score | Query | Match | Length | DB ID    | Description  |
|------------|-------|-------|-------|--------|----------|--|
| 1          | 231   | 100.0 | 231   | 21     | AZ255983 | Core fragment B, a Murine genomic DNA sequence of ne             |
| 2          | 231   | 100.0 | 393   | 21     | AZ255986 | Nucleotide sequence of ne  |
| 3          | 224   | 97.0  | 226   | 21     | AZ255987 | Nucleotide sequence of ne  |
| 4          | 114.4 | 49.5  | 75798 | 25     | ABX77212 | DNA probe 3 detect   |
| c          | 89.4  | 38.7  | 18910 | 25     | ABZ80240 | Human histone d2c  |
| 5          | 89.4  | 38.7  | 5018  | 6      | AAM50150 | Human ovarian cancer   |
| 6          | 89.4  | 38.7  | 5018  | 8      | AAN70646 | Human Colon cancer   |
| 7          | 89.4  | 38.7  | 3336  | 6      | AAN50149 | Murine Post translational regulatory element; PRB; murine; NCTE; |
| 8          | 87.8  | 38.0  | 30    | 21     | AAS55984 | Post-transcriptional RNA nucleo-cytoplasmic transport element;   |

|  |          |
|--|----------|
| Nucleotide sequence of ne  | AAV04900 |
| Nucleotide sequence of ne  | ABY11082 |
| DNA sequence of ne   | AAV04901 |
| DNA sequence of ne   | ABY11083 |
| DNA probe 3 detect   | ABR85497 |
| Human histone d2c  | AAC89560 |
| Human ovarian cancer   | ABP81597 |
| Human Colon cancer   | AA01644  |
| Murine Post translational regulatory element; PRB; murine; NCTE; | AAC46566 |
| Novel human DNA repair gene                                      | AAS32249 |
| EST with homology to human fibrillin 3                           | AB867552 |
| Expressed sequence tag   | AAS08623 |
| Human full-length cDNA   | AAS35795 |
| Human fibrillin secreted protein                                 | AAS46634 |
| Human fibrillin 3  | AAC69907 |
| Human fibrillin secreted protein                                 | AAC46562 |
| Human fibrillin 3  | ABU40559 |
| DNA encoding human fibrillin 3                                   | ABT71695 |
| DNA encoding human fibrillin 3                                   | ABT71693 |
| DNA encoding human fibrillin 3                                   | ABT71691 |
| DNA encoding human fibrillin 3                                   | ABT71690 |
| DNA encoding human fibrillin 3                                   | ABT71692 |
| DNA encoding human fibrillin 3                                   | ABT83579 |
| Arabidopsis thaliana fibrillin 3                                 | AAQ51239 |
| Arabidopsis thaliana fibrillin 3                                 | AAQ35135 |
| DNA encoding human fibrillin 3                                   | AAQ33056 |
| DNA encoding human fibrillin 3                                   | AAQ33280 |
| Human Colon cancer   | AAQ01665 |
| Genomic sequence #   | ABF42209 |
| P. chrysogenum subsp. chrysogenum                                | AAZ29467 |
| Genomic P. chrysogenum   | AAZ29469 |
| Sequence encoding Drosophila melanogaster                        | AAQ51114 |
| Drosophila melanogaster  | ABD07213 |
| Drosophila melanogaster  | ABD07212 |
| Human DNA sequence   | AAS94790 |

## ALIGNMENTS

|          |   |
|----------|---|
| RESULT 1 | AZ255983 standard; DNA; 231 BP.   |
| ID       | AZ255983  |
| AC       | AZ255983;   |
| XX       | XX  |
| XX       | XX  |
| XX       | XX  |
| DT       | 10-APR-2000 (first entry)   |
| DE       | Core fragment B, a murine genomic core PRE region.  |
| XX       | Post translational regulatory element; PRB; murine; NCTE;   |
| KW       | post-transcriptional RNA nucleo-cytoplasmic transport element;  |
| KW       | virus; retrovirus; HIV-1; intracisternal A-type particle; IAP; vaccine; viral infection; core fragment B; ds. |
| XX       | Mus sp.   |
| XX       | WC9961596-A2.   |
| PD       | 02-DEC-1999.  |
| XX       | OS  |
| PP       | 18-MAY-1999; 99WO-US11082.  |
| PR       | 22-MAY-1998; 98US-0086487.  |
| PA       | (USSH ) US DEPT HEALTH & HUMAN SERVICES.  |
| XX       | Pavlakis GN Nappi F;  |
| DR       | WPI; 2000-116336/10.  |
| XX       | Sequence of the cd  |

PT Novel post-transcriptional regulatory elements used to construct  
 PT attenuated retroviruses for vaccines -  
 PS Claim 3; Page 44; 59pp; English.

XX This sequence represents a core post-transcriptional regulatory  
 CC element (PRE) region, designated core fragment B, which is located  
 CC within PRE7 (AAZ55986). Core fragment B can function as a post-  
 transcriptional RNA nucleo-cytoplasmic transport element (NCTE). The PRE  
 CC is used to replace the NCTE of HIV-1, producing a virus with lower  
 replicative activity, resulting in an attenuated virus. This novel PRE  
 CC was initially derived from a murine genomic sequence and sequence  
 CC analysis found that it had significant homology to intracisternal A-type  
 CC particle (IAP) sequences. The nucleic acids and oligonucleotides of the  
 CC invention can be delivered into cells, cultures, tissues and organisms  
 CC for synthesis, mutation and screening. When the PRE is used to replace  
 CC the NCTE of viruses, especially retroviruses such as HIV, an attenuated  
 CC virus is produced, which may be used in viral vaccine for the  
 CC prophylaxis or amelioration of a viral infection in a mammal.  
 XX Sequence 231 BP; 47 A; 58 C; 70 G; 56 T; 0 other;

Query Match 100.0%; Score 231; DB 21; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-70; Indels 0; Gaps 0;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 231; DB 21; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.e-70;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGGGGTGCAGGGCTAAGGCACTGGCACAGGGATAGCTTGCTGGCATCCTGTGGAAAG 60  
 Db 1 GTGGGGTGCAGGGCTAAGGCACTGGCACAGGGATAGCTTGCTGGCATCCTGTGGAAAG 60  
 QY 61 CACGCTGATTGATGAAGGTTCAAGTGTCTAGTCCPCTCCCCAGGAAAAAGACAG 120  
 Db 61 CACGCTGATTGATGAAGGTTCAAGTGTCTAGTCCPCTCCCCAGGAAAAAGACAG 120  
 QY 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGCTTAGGGCCCT 180  
 Db 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGCTTAGGGCCCT 180  
 QY 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 Db 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 QY 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 Db 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 RESULT 2  
 ID AAZ55986 standard; DNA; 393 BP.  
 XX  
 AC AAZ55986;  
 XX  
 DT 10-APR-2000 (first entry)  
 DE Murine genomic post-transcriptional regulatory element, PRE7.  
 XX  
 KW Post translational regulatory element; PRE; murine; NCTE;  
 KW post-transcriptional RNA nucleo-cytoplasmic transport element;  
 KW attenuated virus; retrovirus; HIV-1; intracisternal A-type particle; IAP;  
 KW vaccine; viral infection; PRE7; ds.  
 OS Mus sp.  
 XX  
 PN W09961596-A2.  
 PD 02-DEC-1999.  
 XX  
 PF 18-MAY-1999; 99WO-US1082.  
 XX  
 PR 22-MAY-1998; 98US-0086487.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Pavlakis GN, Napoli F;  
 XX  
 DR 2000-116336/10.

XX Novel post-transcriptional regulatory elements used to construct  
 PT attenuated retroviruses for vaccines -  
 PS Example 1; Page 44; 59pp; English.

XX The invention relates to the core region of a novel post-transcriptional  
 CC regulatory element (PRE), designated core fragment B (AAZ55987), that  
 CC can function as a post-transcriptional RNA nucleo-cytoplasmic transport  
 CC element (NCTE). The PRE is used to replace the NCTE of HIV-1, producing  
 CC a virus with lower replicative activity, resulting in an attenuated  
 CC virus. This novel PRE was initially derived from a murine genomic  
 CC sequence and sequence analysis found that it had significant homology to  
 CC intracisternal A-type particle (IAP) sequences. The nucleic acids and  
 CC oligonucleotides of the invention can be delivered into cells, cultures,  
 CC tissues and organisms for synthesis, mutation and screening. When the  
 CC PRE is used to replace the NCTE of viruses, especially retroviruses such  
 CC as HIV, an attenuated virus is produced, which may be used as a viral  
 CC vaccine for the prophylaxis or amelioration of a viral infection in a  
 CC mammal. This sequence represents PRE7, a murine sequence which comprises  
 CC the core fragment B which was characterised in an exemplification of the  
 CC present invention, and is functional as a PRE.  
 XX Sequence 393 BP; 85 A; 97 C; 109 G; 102 T; 0 other;  
 SQ Query Match 100.0%; Score 231; DB 21; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.e-70;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGGGGTGCAGGGCTAAGGCACTGGCACAGGGATAGCTTGCTGGCATCCTGTGGAAAG 60  
 Db 99 GTGGGGTGCAGGGCTAAGGCACTGGCACAGGGATAGCTTGCTGGCATCCTGTGGAAAG 158  
 QY 61 CACGCTGATTGATGAAGGTTCAAGTGTCTAGTCCPCTCCCCAGGAAAAAGACAG 120  
 Db 61 CACGCTGATTGATGAAGGTTCAAGTGTCTAGTCCPCTCCCCAGGAAAAAGACAG 120  
 QY 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGCTTAGGGCCCT 180  
 Db 121 GGAGCTGCCAAGAACCTCTGGGTGATGAGCTTAAGGGATGGTTTGCTTAGGGCCCT 180  
 QY 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 Db 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 QY 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 Db 181 ATGCCTGACACTGGGATCAGACCTCTACCTTCACCCATGAGSCTTGCTT 231  
 RESULT 3  
 ID AAZ55987  
 XX  
 AC AAZ55987;  
 XX  
 DT 10-APR-2000 (first entry)  
 DE Post-transcriptional regulatory element PRE7 functional fragment, M4.  
 XX  
 KW Post translational regulatory element; PRE; murine; NCTE;  
 KW post-transcriptional RNA nucleo-cytoplasmic transport element; IAP;  
 KW attenuated virus; retrovirus; HIV-1; intracisternal A-type particle; IAP;  
 KW vaccine; viral infection; M4; ds.  
 OS Unidentified.  
 XX  
 PN W0961596-A2.  
 PD 02-DEC-1999.  
 XX  
 PF 18-MAY-1999; 99WO-US1082.  
 XX  
 PR 22-MAY-1998; 98US-0086487.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX

(USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX

|                       |   |            |  |
|-----------------------|---|------------|--|
| PI                    | Pavlakis GN, Nappi F;   | PR         | 18-APR-2001; 2001AU-0004467.   |
| XX                    |   | XX         |  |
| DR                    | WPI; 2000-116336/10.  | PA         | (GENE-) GENE STREAM PTY LTD.   |
| XX                    |   | XX         |  |
| PT                    | Novel post-transcriptional regulatory elements used to construct  | PI         | Daly JM;   |
| PT                    | attenuated retroviruses for vaccines -  | XX         | WPI; 2003-093021/08.   |
| XX                    |   | XX         |  |
| PS                    | Example 1; Page 45; 59pp; English.  | PT         | New transgenic non-human animal expressing a foreign polypeptide associated with drug behavior and/or metabolism, useful for studying  |
| XX                    |   | PT         | the behavior and/or metabolism of a drug in other animals -  |
| CC                    | The invention relates to the core region of a novel post-transcriptional regulatory element (PRE), designated core fragment B (AA555983), that can function as a post-transcriptional RNA nucleo-cytoplasmic transport element (NCTE). The PRE is used to replace the NCTE of HIV-1, producing a virus with lower replicative activity, resulting in an attenuated virus. This novel PRE was initially derived from a murine genomic sequence and sequence analysis found that it had significant homology to intracisternal A-type particle (IAP) sequences. The nucleic acids and oligonucleotides of the invention can be delivered into cells cultures, tissues and organisms for synthesis, mutation and screening. When the PRE is used to replace the NCTE of viruses, especially retroviruses such as HIV, an attenuated virus is produced, which may be used in a viral vaccine for the prophylaxis or amelioration of a viral infection in a mammal. This sequence represents M4, a sequence which is very similar to core fragment B (having 7 additional nucleotides at the 3' end and 2 fewer nucleotides at the 5' end). M4 was characterised in an exemplification of the present invention, and is functional as a PRE. | XX         | This invention relates to a transgenic non-human animal which may be used for assessing the behaviour and/or metabolism of a drug in another animal and which expresses a foreign polypeptide associated with drug behaviour and/or metabolism. The invention also comprises a nucleic acid construct for use in producing the above transgenic non-human animal and a method of assessing the metabolism and/or behavior of a drug in an animal of interest, comprising administering a test agent to the transgenic animal and conducting analytical tests to determine drug metabolism and/or behaviour. The transgenic animal is useful in studying drug metabolism and/or behaviour in other animals. The nucleic acid construct is useful in producing the above transgenic animal and the methods are used for producing, breeding and using transgenic animals for pharmacological (e.g., pharmacodynamic or pharmacokinetic assays) and/or toxicological studies. Nucleic acid sequences used within the invention are serum albumin; alpha-acidic glycoprotein; cytochrome P450 (CYP); uridine diphosphoglucuronosyl transferase (UGT); multidrug resistance proteins and (MRP's). The present sequence represents a CC primer used to create a transgenic animal within the scope of the invention. |
| SO                    | Sequence 226 BP; 47 A; 58 C; 69 G; 52 T; 0 other;   | SQ         | Sequence 75798 BP; 21534 A; 17028 C; 16619 G; 20617 T; 0 other;  |
| Query Match           | 97.0%   | Score      | 224;   |
| Best Local Similarity | 100.0%  | DB         | 21;  |
| Matches               | 224;  | Length     | 226;   |
| Conservative          | 0;  | Pred. No.  | 8e-68;   |
| Mismatches            | 0;  | Mismatches | 0;   |
| Indels                | 0;  | Gaps       | 0;   |
| Db                    |   | Db         |  |
| Qy                    | 1 GTGGGTGCGAGGTAAAGACTGACAGAGGATAGCTTGATGCTGTGGAAAG 60  | Qy         | 1 GTGGGTGCGAGGTAAAGCTTCAGTTCTAGTCAGTGCCTTCAGGAAAAACGACAG 60  |
| Db                    | 3 GTGGGTGCGAGGTAAAGACTGACAGAGGATAGCTTGATGCTGTGGAAAG 62  | Db         | 3 CACGHTCATTTGATTAAGGGATGGTTTGTAGGGCCCT 120  |
| Qy                    | 61 CACGHTCATTTGATTAAGGGATGGTTTGTAGGGCCCT 120  | Qy         | 61 CACGHTCATTTGATTAAGGGATGGTTTGTAGGGCCCT 120   |
| Db                    | 63 CACGHTCATTTGATTAAGGGATGGTTTGTAGGGCCCT 122  | Db         | 63 CACGHTCATTTGATTAAGGGATGGTTTGTAGGGCCCT 122   |
| Qy                    | 121 GGAGCTGCCAAGACCTCTGGGCTAGGATGAGCTTCTGGTACGGCCCT 180   | Qy         | 121 GGAGCTGCCAAGACCTCTGGGCTAGGATGAGCTTCTGGTACGGCCCT 180  |
| Db                    | 123 GGAGCTGCCAAGACCTCTGGGCTAGGATGAGCTTCTGGTACGGCCCT 182   | Db         | 123 GGAGCTGCCAAGACCTCTGGGCTAGGATGAGCTTCTGGTACGGCCCT 182  |
| Qy                    | 181 ATGCTTGCACTGGGATGAGCTTACCCATGAGG 224  | Qy         | 181 ATGCTTGCACTGGGATGAGCTTACCCATGAGG 224   |
| Db                    | 183 ATGCTTGCACTGGGATGAGCTTACCCATGAGG 226  | Db         | 183 ATGCTTGCACTGGGATGAGCTTACCCATGAGG 226   |
| RESULT 4              |   | RESULT 5   |  |
| ID                    | ABX77212 standard; DNA; 75798 BP.   | ID         | ABZ0240/c  |
| XX                    |   | XX         |  |
| AC                    | ABX77212;   | AC         | ABZ0240 standard; cDNA; 1870 BP.   |
| XX                    |   | XX         |  |
| DT                    | 01-MAY-2003 (first entry)   | AC         | ABZ0240;   |
| XX                    |   | XX         |  |
| DE                    | Mouse uridine diphosphoglucuronosyl transferase gene locus.   | DT         | 02-JUN-2003 (first entry)  |
| XX                    |   | XX         |  |
| KW                    | Human; ss; transgenic; drug metabolism; behaviour; PCR; primer; mouse; pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin; alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP; uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.   | DE         | Mouse tramordin 3' RACE product encoding cDNA SEQ ID NO:40.  |
| KW                    | Homo sapiens.   | XX         | Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy; central nervous system disorder; CNS disorder; multiple sclerosis; nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramodin; mouse; tramordin 3'; gene; ss.  |
| OS                    | PN WO200283897-A1.  | XX         |  |
| XX                    |   | OS         |  |
| PD                    | 24-OCT-2002.  | XX         |  |
| XX                    |   | FH         |  |
| PF                    | 18-APR-2002; 2002WO-AU00485.  | FT         |  |
| XX                    |   | FT         |  |
|                       |   |            | /tag= a  |

FT /partial  
 FT /product= "tramordin 3 fragment"  
 FT /note= "no start codon given"  
 XX WO2003016502-A2.  
 XX 27-FEB-2003.  
 XX 21-AUG-2002; 2002WO-US26637.  
 XX PR 21-AUG-2001; 2001US-313907.  
 XX PR 21-AUG-2002; 2002US-022510.  
 PA (MCCLA ) MC LAUGHLIN RES INST.  
 XX PI Birmingham JR;  
 XX DR WPI; 2003-278567/27.  
 DR P-PSDB; ABP96445.  
 XX PR New nucleic acid sequence encoding tramordin, e.g. mouse tramd 1, mouse tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1 or rat tramd 3 useful for treating CNS, e.g. stroke, multiple sclerosis, trauma, neuropathic pain  
 PR PS Example 11; Fig 23C; 177pp; English.  
 XX CC The present invention describes an isolated nucleic acid sequence comprising a cDNA sequence encoding mouse tramordin (tramd) 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is located to chromosome 11, whereas human tramd 1 is located to chromosome 5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic and cerebroprotective activities, and can be used in gene therapy. The nucleic acid sequences are useful for diagnosing and treating central nervous system (CNS) disorders such as multiple sclerosis, nerve injury, neuropathic pain, stroke or trauma, and non-CNS disorders. The present sequence encodes a mouse tram d 3, RACE product, which is given in the exemplification of the present invention.  
 XX SQ Sequence 1870 BP; 483 A; 484 C; 429 G; 474 T; 0 other;  
 Query Match 38.7%; Score 89.4; DB 25; Length 1870;  
 Best Local Similarity 74.9%; Pred. No. 1.1e-20;  
 Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;  
 Qy 55 GGAAGGCCAGCTGTGATGGTCAAGTAAAGCTTCAGTCCTCCAGAAAC 114  
 Db 1347 GGAGAACATGTCTCATTTCTAGAAGTTCAAGTGTCAAGTCTCCCTCCAGGAAAC 1288  
 Qy 115 GACACGGGAGGTGGCCAAGACCTCTGGTGA-----TGAGGCTTAAGGGATGGTT 167  
 Db 1287 GACACGGGAGGTGGCCAAGCTTAAGGCTTAAGGCTTAATCCCT 1228  
 Qy 168 GTGTAGGCCCTATGCTGCACATGGGATCAGACCTTACCTCACCC 218  
 Db 1227 GTACATGGCCTTAACTACACAGGGATTGACCTCTATCCACTC 1177  
 RESULT 6  
 AAN50150  
 ID AAN50150 standard; cDNA; 5018 BP.  
 AC AAN50150;  
 XX DE Sequence of the cDNA clone 23B6p10.2 encoding a polypeptide exhibiting mammalian immunoglobulin binding factor activity (IBF).  
 DE DE Antibody; immunoglobulin; ss.  
 XX KW Unidentified.

KW immunoglobulin E-binding factor; ss.  
 XX Rattus sp.  
 OS XX Key  
 FN FH CDS Location/Qualifiers 423..2750  
 PD FT /\*tag= a  
 XX PN EP155192-A.  
 XX PR 16-MAR-1984;  
 XX PR 84US-0590430.  
 PA (SCHE ) SCHERRING BIOTECH CORP.  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 - PA (SCHE ) SCHERRING-BIOTECH CO.  
 PA (DNAX ) DNAX RES INST MOLEC.  
 XX PI Martens CL, Ishizaka K, Moore KW, Huff TF;  
 XX DR WPI; 1985-231863/38.  
 DR P-PSDB; AAP50121.  
 XX PT New complementary DNA clones coding for poly(peptide(s)) with PT sequence of mammalian immunoglobulin factor and obtid. from PT transformed or transfected host  
 XX PS Claim 8; Page 59-64; 71pp; English.  
 XX CC IBF and IBF cDNA are useful in studies on the immune system. Treatment for IgE-mediated diseases may be possible. IBF may enhance B-cell differentiation into an immunoglobulin-secreting cell. (Updated on 16-AUG-2002 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX SQ Sequence 5018 BP; 1424 A; 1086 C; 1165 G; 1343 T; 0 other;  
 Query Match 38.7%; Score 89.4; DB 6; Length 5018;  
 Best Local Similarity 74.9%; Pred. No. 1.6e-20;  
 Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;  
 Qy 55 GGAAGGCCAGCTGTGATGGTCAAGTAAAGCTTCAGTCCTCCAGAAAC 114  
 Db 4449 GGAGAACATGTCTCATTTCTAGAAGTTCAAGTGTCAAGTCTCCAGGAAAC 4508  
 Qy 115 GACACGGGAGGTGGCCAAGACCTCTGGTGA-----TGAGGCTTAAGGGATGGTT 167  
 Db 4509 GACACGGGAGGTGGCCAAGACCTCTGGTGA-----TGAGGCTTAAGGGATGGTT 167  
 Qy 168 GTGTAGGCCCTATGCTGCACATGGGATCAGACCTTACCTCACCC 218  
 Db 4569 GTACATGGCCTTAACTACACAGGGATTGACCTCTATCCACTC 4619  
 RESULT 7  
 AAN70646  
 ID AAN70646 standard; cDNA; 5018 BP.  
 XX AC AAN70646;  
 XX DE 25-MAR-2003 (updated)  
 DT 10-MAR-2003 (updated)  
 DT 27-APR-1991 (first entry)  
 DE Clone 23B6p102 encoding polypeptide with IgE binding factor activity.  
 DE KW Antibody; immunoglobulin; ss.  
 XX OS Unidentified.





Db 904 GACATGTCATCTTCATGAGGTTCAGTCTCCTAGTCCAGTCAGCAANNAAC 963  
 CC pathogen. These therapeutic products can be used in gene therapy. The  
 CC artificial chromosomes are useful as cloning vehicles that accommodate  
 CC entire genomes for the preparation of genomic DNA libraries, and also  
 CC for the production of proteins which may be involved in a biochemical  
 CC pathway or in multivalent vaccines.  
 XX  
 SQ Sequence 1400 BP; 341 A; 310 C; 345 G; 404 T; 0 other;  
 Query Match 35.2%; Score 81.4%; DB 18; Length 1400;  
 Best Local Similarity 71.9%; Pred. No. 6.3e-18;  
 Matches 123; Conservatism 0; Mismatches 41; Indels 7; Gaps 1;  
 Qy 55 GGAGGGCACGCTCATGGATGAGGTTCAGTGCCTAGTTCCCTCCCCAGGAAAAAC 114  
 Db 771 GGAGACAGTCATCTTCAGCAGTCAGTGCCTTCAGCAGAAC 830  
 AC 115 GACACGGGAGCTGGCCAAAGCCCTCTGGGTGA-----TGAGCCTAAGGGATGGTTTT 167  
 DT 831 GACACGGGAGCTGGCCAAAGGGTGTGAGCTTAAGGGCTAATGCT 890  
 XX Nucleotide sequence of a portion of a neo-minichromosome.  
 DE 168 GTGTAGGGGCCATTGCTGCAACTGGGATCAGACCTCTACATTCAACC 218  
 KW Neo-minichromosome; multicentric; dicentric; chromosome; euchromatin;  
 KW heterochromatic DNA; minichromosome; artificial chromosome;  
 KW selectable marker; satellite artificial chromosome; SATAC; genomic locus;  
 KW targeted integration; transgenic animal; therapeutic product;  
 KW gene therapy; cloning vehicle; genomic DNA library; ss.  
 KW RESULT 12  
 ABX11083 ID ABX11083 standard; DNA; 1400 BP.  
 XX  
 AC ABX11083;  
 PN 24-APR-2003 (first entry)  
 XX  
 DE DNA sequence of neo-minichromosome PCR product #2 from EC3/7C5 cells.  
 XX  
 KW Nucleic acid amplification; artificial chromosome isolation; MAC;  
 KW DNA delivery; mammalian artificial chromosome; gene therapy; organ;  
 KW humanised genetically transformed animal; chromosomal element;  
 KW gene product production system; transgenic; centromere function;  
 KW information storage vehicle; artificial chromosome vector; human;  
 KW species-specific artificial chromosome; mouse; bacteriophage lambda;  
 KW bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; mutant; ds.  
 XX  
 DE ABX11083  
 XX  
 AC ABX11083;  
 DT 24-APR-2003 (first entry)  
 XX  
 DE DNA sequence of neo-minichromosome PCR product #2 from EC3/7C5 cells.  
 XX  
 KW Nucleic acid amplification; artificial chromosome isolation; MAC;  
 KW DNA delivery; mammalian artificial chromosome; gene therapy; organ;  
 KW humanised genetically transformed animal; chromosomal element;  
 KW gene product production system; transgenic; centromere function;  
 KW information storage vehicle; artificial chromosome vector; human;  
 KW species-specific artificial chromosome; mouse; bacteriophage lambda;  
 KW bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; mutant; ds.  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Bacteriophage lambda.  
 OS Bacteriaceae.  
 OS Synthetic.  
 XX  
 FH Key misc\_feature 1..1400  
 FT \*tag= a  
 FT Location/Qualifiers  
 /note= "All N nucleotides have been added by the  
 indexer to match the number of bases this  
 sequence is listed to contain in the  
 sequence listing."  
 XX  
 PR 07-NOV-2000; 200005-0724693.  
 PR 10-APR-1997; 97US-0815682.  
 PR 10-APR-1996; 96US-0639822.  
 PR 15-JUL-1996; 96US-0662080.  
 XX  
 PA (AMGE-) AMERICAN GENE THERAPY INC.  
 PA (BIOL-) BIOLOGICAL RES CENT HUNGARIAN ACAD SCI.  
 PA (UTLO-) UNIV LOMA LINDA.  
 XX  
 PI Hadlaczky G, Szalay AA;  
 XX  
 DR WPI; 1997-535860/49.  
 XX  
 PT Producing satellite artificial chromosomes or mini:chromosomes -  
 CC useful for, e.g. cloning multiple proteins of a metabolic pathway or  
 CC multivalent vaccines, etc.  
 XX  
 PS Claim 85; Page 204; 248pp; English.  
 XX  
 CC Nucleotide sequences AAV04900-902 are portions of a neo-minichromosome  
 CC isolated from a mouse rRNA gene. 3 products were obtained from a single  
 CC amplification reaction, which suggests that the sequence of the DNA  
 CC located between different sets of inverted repeats may differ.  
 CC  
 CC AAV04900-01 show high (96%) sequence homology to portions of DNA from  
 CC intracisternal A-particles from mouse. The minichromosome is derived from  
 CC multicentric, typically dicentric, chromosome. The minichromosome contains more  
 CC euchromatin than heterochromatic DNA. The minichromosome is an example  
 CC of an artificial chromosome. Artificial chromosomes can be produced by  
 CC incorporating a DNA fragment comprising a selectable marker into a cells'  
 CC genomic DNA, into or adjacent to an amplifiable region, and selecting a  
 CC cell that comprises either a satellite artificial chromosome (SATAC)  
 CC (see AAV04903-09), a de novo centromere, or a minichromosome (e.g.  
 CC present sequence). Artificial chromosomes provide an extra genomic locus  
 CC for targeted integration of megabase size DNA fragments that contain  
 CC single or multiple genes. SATACs can be introduced into embryonic cells  
 CC of non-human animals to produce transgenic animals that express a  
 CC heterologous DNA encoding a therapeutic product, e.g. anti-HIV. The DNA  
 CC may also encode antigens that induce immunoprotective response against

|    |  |  |    |  |
|----|--|--|----|--|
| XX | WPI; 2003-265757/26.   | PR; 15-JUL-1993; 93DE-4323727.   | XX | PR; 15-JUL-1993; 93DE-4323727.   |
| XX | Amplifying nucleic acid for constructing artificial chromosomes, comprising introducing a DNA fragment into a cell to incorporate the DNA into a chromosome, and identifying cells with chromosomes that has undergone amplification -   | PT (BOEFL ) BOEHRINGER MANNHEIM GMBH.  | XX | PA (BOEFL ) BOEHRINGER MANNHEIM GMBH.  |
| XX | Example 2; Page 64; 107PP; English..   | PI Abken H, Albert W, Jungfer H, Abken HJ;   | XX | PI Abken H, Albert W, Jungfer H, Abken HJ;   |
| PS |  | DR; 1995-067344/09.  | XX | DR; 1995-067344/09.  |
| XX | The present invention relates to methods for amplifying nucleic acids in cells, for isolating artificial chromosomes, and preparing cell lines that contain artificial chromosomes. The methods comprise introducing a DNA fragment comprising a selectable marker into a cell, growing the cell under selective conditions to produce cells that have incorporated the DNA fragment or its portion into a chromosome, and identifying from among the resulting cells those that include a chromosome or its fragment with portion that has undergone amplification. The invention also discloses cell lines that contain the nucleic acids or artificial chromosomes. The methods of the invention are useful for amplifying nucleic acids in cells, in generating and isolating artificial chromosomes (e.g. mammalian artificial chromosomes or MACs), and in delivering the chromosomes to selected cells and tissues. The artificial chromosomes are useful in gene therapy, gene product production systems, production of transgenic plants and animals that would employ chromosomal elements as information storage vehicles, for analysis and study of centromere function, for the production of artificial chromosome vectors, and for the preparation of species-specific artificial chromosomes. The present sequence represents a neo-minichromosome (composed of human, bacteriophage lambda, plasmid, bacterial and mouse DNAs) PCR product from EC3/7C5 cells that is used in the method of the present invention. | CC   | CC |  |
| XX | Sequence 1400 BP; 330 A; 299 C; 331 G; 388 T; 52 other;  | Query Match 34.0%; Score 78.6; DB 25; Length 1400; Best Local Similarity 69.6%; Pred. No. 5.9e-17; Matches 119; Conservative 0; Mismatches 45; Indels 7; Gaps 1;   | XX | Query Match 15.2%; Score 35; DB 16; Length 50; Best Local Similarity 100.0%; Pred. No. 0.027; Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |
| SO |  | Sequence 50 BP; 13 A; 15 C; 9 G; 13 T; 0 other;  | XX | Sequence 50 BP; 13 A; 15 C; 9 G; 13 T; 0 other;  |
| XX |  | Query Match 15.2%; Score 35; DB 16; Length 50; Best Local Similarity 100.0%; Pred. No. 0.027; Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  | XX | Query Match 15.2%; Score 35; DB 16; Length 50; Best Local Similarity 100.0%; Pred. No. 0.027; Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |
| QY | 197 GATCAGACCTCTAACCTTCAACCATGAGCTGTGCTT 231   | XX   | XX | XX   |
| Db | 1 GATCAGACCTCTAACCTTCAACCATGAGCTGTGCTT 35  | XX   | XX | XX   |
| XX |  | RESULT 14 AAC89560/C   | XX | RESULT 14 AAC89560/C   |
| XX |  | ID AAC89560 standard; DNA: 122186 BP.  | XX | ID AAC89560 standard; DNA: 122186 BP.  |
| XX |  | AC AAC89560;   | XX | AC AAC89560;   |
| XX |  | DT 08-MAR-2001 (first entry)   | XX | DT 08-MAR-2001 (first entry)   |
| XX |  | DE Human histone deacetylase HDAC-D coding sequence.   | XX | DE Human histone deacetylase HDAC-D coding sequence.   |
| XX |  | KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;  | XX | KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;  |
| XX |  | KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense; ds.  | XX | KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense; ds.  |
| OS |  | XX Homo sapiens.   | XX | XX Homo sapiens.   |
| XX |  | PN WO2000071703-A2.  | XX | PN WO2000071703-A2.  |
| XX |  | XX 30-NOV-2000.  | XX | XX 30-NOV-2000.  |
| XX |  | PD   | XX | PD   |
| XX |  | PF 03-MAY-2000; 2000WO-IB01252.  | XX | PF 03-MAY-2000; 2000WO-IB01252.  |
| XX |  | AC 03-MAY-1999; 99US-0132287.  | XX | AC 03-MAY-1999; 99US-0132287.  |
| XX |  | PA (METH-) METHYLENE INC.  | XX | PA (METH-) METHYLENE INC.  |
| XX |  | PI Macleod AR, Li Z, Beasterman JM;  | XX | PI Macleod AR, Li Z, Beasterman JM;  |
| XX |  | XX DR WO2001-016407/02.  | XX | XX DR WO2001-016407/02.  |
| XX |  | XX Antisense oligonucleotide that inhibits expression of a histone deacetylase, useful for treating and/or alleviating the symptoms of neoplasia, or for inhibiting neoplastic cell growth in an animal -                    | CC | XX Antisense oligonucleotide that inhibits expression of a histone deacetylase, useful for treating and/or alleviating the symptoms of neoplasia, or for inhibiting neoplastic cell growth in an animal -                    |
| XX |  | PS Disclosure; Page 89-125; 125PP; English.  | CC | PS Disclosure; Page 89-125; 125PP; English.  |
| XX |  | XX The present invention provides inhibitors of histone deacetylase enzymes such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These inhibitors may be antisense strands or they may be compounds identified | CC | XX The present invention provides inhibitors of histone deacetylase enzymes such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These inhibitors may be antisense strands or they may be compounds identified |



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 21:30:48 ; Search time 1470 Seconds  
(without alignments)  
64.28.656 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 gttgggtgcaggtaagca.....ttcacccataggcttgctt 231

Scoring table: IDENTITY\_NUC

Gapon 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

| Result No. | Score | Query % | Match % | Length | DB ID        | Description                   |
|------------|-------|---------|---------|--------|--------------|-------------------------------|
| 1          | 231   | 100.0   | 100.0   | 1086   | 10 AF250998  | AF250998 Mus muscu            |
| 2          | 231   | 100.0   | 100.0   | 1086   | 10 BC200799  | BC200799 Mus muscu            |
| 3          | 231   | 100.0   | 100.0   | 2701   | 10 AC10548   | AC10548 Mus muscu             |
| 4          | 231   | 100.0   | 100.0   | 42511  | 2 AC10548    | AC10548 Mus muscu             |
| 5          | 231   | 100.0   | 100.0   | 64042  | 2 AC10424    | AC10424 Mus muscu             |
| 6          | 231   | 100.0   | 100.0   | 66393  | 2 AC10424    | AC10424 Mus muscu             |
| 7          | 231   | 100.0   | 100.0   | 66489  | 2 AC104278   | AC104278 Mus muscu            |
| 8          | 231   | 100.0   | 100.0   | 66489  | 2 AC104278   | AC104278 Mus muscu            |
| 9          | 231   | 100.0   | 100.0   | 70360  | 2 AC101122   | AC101122 Mus muscu            |
| 10         | 231   | 100.0   | 100.0   | 71109  | 10 AL935271  | AL935271 Mouse DNA            |
| 11         | 231   | 100.0   | 100.0   | 77127  | 10 BX294125  | BX294125 Mouse DNA            |
| 12         | 231   | 100.0   | 100.0   | 84250  | 10 AF481949  | AF481949 Mus muscu            |
| 13         | 231   | 100.0   | 100.0   | 98653  | 10 AL731664  | AL731664 Mouse DNA            |
| 14         | 231   | 100.0   | 100.0   | 110000 | 2 AC12401_0  | AC12401_0 Mus muscu           |
| 15         | 231   | 100.0   | 100.0   | 110000 | 2 AL939223_2 | AL939223_2 Continuation (3 of |
| 16         | 231   | 100.0   | 100.0   | 132457 | 2 AC102180   | AC102180 Mus muscu            |
| 17         | 231   | 100.0   | 100.0   | 134079 | 2 AC17697    | AC17697 Mus muscu             |
| 18         | 231   | 100.0   | 100.0   | 134264 | 10 BX005191  | BX005191 Mouse DNA            |
| 19         | 231   | 100.0   | 100.0   | 134717 | 10 AL626805  | AL626805 Mouse DNA            |
| 20         | 231   | 100.0   | 100.0   | 141025 | 10 AL807250  | AL807250 Mouse DNA            |
| 21         | 231   | 100.0   | 100.0   | 142902 | 2 AC119234   | AC119234 Mus muscu            |
| 22         | 231   | 100.0   | 100.0   | 14387  | 10 AL844603  | AL844603 Mouse DNA            |
| 23         | 231   | 100.0   | 100.0   | 149172 | 10 AC127278  | AC127278 Mus muscu            |
| 24         | 231   | 100.0   | 100.0   | 150017 | 2 AL773509   | AL773509 Mus muscu            |
| 25         | 231   | 100.0   | 100.0   | 155071 | 2 AC137713   | AC137713 Mus muscu            |
| 26         | 231   | 100.0   | 100.0   | 156256 | 2 AC140843   | AC140843 Mus muscu            |
| 27         | 231   | 100.0   | 100.0   | 161580 | 2 AC118267   | AC118267 Mus muscu            |
| 28         | 231   | 100.0   | 100.0   | 162681 | 10 AC122059  | AC122059 Mus muscu            |
| 29         | 231   | 100.0   | 100.0   | 164939 | 2 AC119839   | AC119839 Mus muscu            |
| 30         | 231   | 100.0   | 100.0   | 165222 | 2 AC121307   | AC121307 Mus muscu            |
| 31         | 231   | 100.0   | 100.0   | 165764 | 2 AC131767   | AC131767 Mus muscu            |
| 32         | 231   | 100.0   | 100.0   | 166764 | 2 AC15875    | AC15875 Mus muscu             |
| 33         | 231   | 100.0   | 100.0   | 166777 | 2 AC124106   | AC124106 Mus muscu            |
| 34         | 231   | 100.0   | 100.0   | 167744 | 10 AC125537  | AC125537 Mus muscu            |
| 35         | 231   | 100.0   | 100.0   | 171826 | 2 AC119810   | AC119810 Mus muscu            |
| 36         | 231   | 100.0   | 100.0   | 172533 | 2 AC120551   | AC120551 Mus muscu            |
| 37         | 231   | 100.0   | 100.0   | 173417 | 10 AL07143   | AL07143 Mouse DNA             |
| 38         | 231   | 100.0   | 100.0   | 175260 | 2 AC138680   | AC138680 Mus muscu            |
| 39         | 231   | 100.0   | 100.0   | 175448 | 2 AC139226   | AC139226 Mus muscu            |
| 40         | 231   | 100.0   | 100.0   | 175692 | 2 AC119918   | AC119918 Mus muscu            |
| 41         | 231   | 100.0   | 100.0   | 176184 | 2 AC122385   | AC122385 Mus muscu            |
| 42         | 231   | 100.0   | 100.0   | 178821 | 10 AC127125  | AC127125 Mouse DNA            |
| 43         | 231   | 100.0   | 100.0   | 179058 | 10 AL27125   | AC137557 Mus muscu            |
| 44         | 231   | 100.0   | 100.0   | 179662 | 2 AC137557   | AC137557 Mus muscu            |
| 45         | 231   | 100.0   | 100.0   | 179999 | 10 BX072541  | BX072541 Mouse DNA            |

## ALIGNMENTS

|            |   |   |                 |
|------------|---|---|-----------------|
| RESULT 1   | AF250998  | 1086 bp DNA linear                          | ROD 16-MAY-2001 |
| LOCUS      | Mus musculus  | RTE-clonetl RNA transport element sequence. |                 |
| DEFINITION |   |   |                 |
| ACCESSION  | AF250998  |   |                 |
| VERSION    | AF250998.1  | G1.14090507                                 |                 |
| KEYWORDS   |   |   |                 |
| SOURCE     | Mus musculus (house mouse)  |   |                 |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus . |   |                 |
| REFERENCE  | Nappi,F., Schneider,R., Zolotukhin,A., Smulevitch,S., Michalowski,D., Bear,J., Felber,B.K. and Pavlakis,G.N.                          |   |                 |
| AUTHORS    | Identification of a novel posttranscriptional regulatory element by   |   |                 |
| TITLE      | Pred. No. is the number of results predicted by chance to have a  |   |                 |

|                          |   |
|--------------------------|---|
| /organism="Mus musculus" |   |
| /mol_type="genomic DNA"  |   |
| /db_xref="taxon:1090"    |   |
| /clone="RTE-clone3"      |   |
| 381 .. 627               |   |
|                          | /note="fragment M1; contains transcriptional regions"   |
| BASE COUNT               | 247 a   |
| ORIGIN                   | 268 c   |
| Query Match              | 100.0%  |
| Best Local Similarity    | 100.0%  |
| Matches                  | 231; Conservative 0; Mismatch 0   |
| Qy                       | 1 GGGGGTCCGAGGTTAAGGACTGACAGAGG   |
| Db                       | 393 GTGGGGTCGAGGCTAAGCAGTGACAGGG  |
| Qy                       | 61 CACGTCTCATGGCTAGGTTCACTGATGCTCT  |
| Db                       | 453 CACGTCTCATGGCTAGGTTCACTGATGCTCT   |
| Qy                       | 121 GGAGCTGCCAAAGACCTCTGGGTGATGAG   |
| Db                       | 513 GGAGCTGCCAAAGACCTCTGGGTGATGAG   |
| Qy                       | 181 ATGCTTGACATCTGGGATCAGACCTCTAAC  |
| Db                       | 573 ATGCTTGACATCTGGGATCAGACCTCTAAC  |
| RESULT                   | 3   |
| BC020078                 | BC020078  |
| LOCUS                    | Mus musculus  |
| DEFINITION               | Similar to RIKEN cDNA MGC_28125 IMAGE_3980327, mRNA, co   |
| ACCESSION                | BC020078  |
| VERSTON                  | BC020078..1   |
| KEYWORDS                 | MGC   |
| SOURCE                   | Mus musculus (house mouse)  |
| ORGANISM                 | Eukaryota; Metazoa; Chordata; Craniata; Eutheria; Rodentia; Sciuroidea; Muridae; Mus  |
| REFERENCE                | 1 (bases 1 to 2701)   |
| AUTHORS                  | Strausberg,R.   |
| TITLE                    | Direct Submission   |
| JOURNAL                  | Submitted (19-DEC-2001) National Gene Collection (NGC), Cancer Genome Institute, 31 Center Drive, Room USA  |
| REMARK                   | NIH MGC Project URL: http://mgc.nigri.nih.gov   |
| COMMENT                  | Contact: MGC help desk Email: cgapbsr@mail.nih.gov Tissue Procurement: Gilbert Smith CDNA Library Preparation: Life Technologies DNA Arrayed by: The I.M. DNA Sequencing by: Baylor College Sequencing Center Center Code: BCM-HGSC Web site: http://www.hgsc.bcm.edu Contact: amg@bcm.tmc.edu Series: TRAK Plate: 36 Row: e Col This clone was selected for full passed the following selection criteria: identity to protein. |
| FEATURES                 | Source 1..2701  |

organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="TAXON:10090"  
/clone="MGC:28125 IMAGE:3980327"  
/tissue\_type="Mammary tumor. Metallothionien-TGF alpha  
model. 10 month old virgin mouse. Taken by biopsy."  
/clone.lib="NCI CGAP\_Mam1"  
/lab\_host="DH10B"  
note="Vector: pCMV-SPORT6"  
/codon\_start=1  
/product="Similar to RIKEN cDNA 1700066C05 gene"  
/protein\_id="AAH20078.1"  
/db\_xref="GI:18043257"  
/clone="MILDQGRQQGSPYQVFKQVNQKIAIRAGSLPQRGEANL  
TLLTPDMOTVIEVNVSSPRGVSTPPGDVQVLMWRSVVCYFPQDQTEPLVPE  
RUMRKCKNEADDPVAVDPTVDDPTSPKDGAEQRDFPGIPPEADTSST"

CDS

| BASE COUNT            | ORIGIN | 744 a  | 542 c         | 789 g        | 626 t |
|-----------------------|--------|--|---------------|--------------|-------|
| Query Match           | 100.0% | Score 231;   | DB 10;        | Length 2701; |       |
| Best Local Similarity | 100.0% | Pred. No. 1..9e-65;  | Mismatches 0; | Gaps 0;      |       |
| Matches               | 231;   | Conservative 0;  | Indels 0;     |              |       |
| Qy                    | 1      | GTGGGTGCGAGGGTAAGCACTGACAGGGATAGCTGCTGCGCATCCTGGAAGG 60          |               |              |       |
| Db                    | 967    | GTGGGTGCGAGGGTAAGCACTGACAGGGATAGCTGCTGCGCATCCTGGAAGG 1026        |               |              |       |
| Qy                    | 61     | CAGCTCTGATTGCCATGAAAGGTTCAAGTGCCTAGTCCTCCCTCCCAGAAAACGACAGC 120  |               |              |       |
| Db                    | 1027   | CACGTCTGATTGCCATGAAAGGTTCAAGTGCCTAGTCCTCCCTCCCAGAAAACGACAGC 1086 |               |              |       |
| Qy                    | 121    | GGACTGGCCAAAGACCTCTGGTAGAGGCTAAGGATGTTTGTTGAGGCCCT 180           |               |              |       |
| Db                    | 1087   | GGACTGGCCAAAGACCTCTGGTAGAGGCTAAGGATGTTTGTTGAGGCCCT 1146          |               |              |       |
| Qy                    | 181    | ATGGTGGCACTGGGATCAGACCTCTACCCCTCACCCCTAGGGATGGCTTCCTT 231        |               |              |       |
| Db                    | 1147   | ATGGTGGCACTGGGATCAGACCTCTACCCCTCACCCCTAGGGCTTCCTT 1197           |               |              |       |

RESULT'4  
AC100548

DEFINITION 42511 bp DNA linear HTG 22-NOV-2001  
LOCUS Mus musculus clone RP23-152017, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC100548  
VERSION 1 GI:17047914  
KEYWORDS HTGS\_PHASE0  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 42511)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Unpublished

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS Anderson, S., Barna, N., Bastien, V., Boguslavkiy, L., Boukhgaiter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J.J., Chazarro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galyan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lebockzy, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McSwain, P., McKernan, P., McSheeters, R., Meidrim, J.,  
Meneus, L., Mihoya, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Schupback, R.,  
Roman, J., Rosetti, M., Roy, A., Santor, R., Schauer, S., Schupback, R.,  
Seaman, S., Sevary, P., Spangler, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talama, J., Testay, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research,  
320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/Rn/RepeatMasker.html  
-----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
-----  
Center Project name: L15671  
Center Clone name: L15671  
Center Clone name: L15671  
\* NOTE: This record contains 54 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 699: contig of 699 bp in length  
\* 700: gap of 100 bp  
\* 800: 1475: contig of 676 bp in length  
\* 1476: 1575: gap of 100 bp  
\* 1576: 2267: contig of 692 bp in length  
\* 2268: 2367: gap of 100 bp  
\* 2368: 3032: contig of 665 bp in length  
\* 3033: 3132: gap of 100 bp  
\* 3133: 3827: contig of 695 bp in length  
\* 3828: 3927: gap of 100 bp  
\* 3928: 4628: contig of 701 bp in length  
\* 4629: 5393: contig of 665 bp in length  
\* 5394: 5493: gap of 100 bp  
\* 5494: 6163: contig of 670 bp in length  
\* 6164: 6263: gap of 100 bp  
\* 6264: 6946: contig of 683 bp in length  
\* 6947: 7046: gap of 100 bp  
\* 7047: 7724: contig of 678 bp in length  
\* 7725: 7824: gap of 100 bp  
\* 7825: 8528: contig of 704 bp in length  
\* 8628: 8626: gap of 100 bp  
\* 8629: 9326: contig of 698 bp in length  
\* 9327: 9426: gap of 100 bp  
\* 10129: 10229: gap of 100 bp  
\* 10230: 10909: contig of 680 bp in length  
\* 10910: 11009: gap of 100 bp  
\* 11010: 11717: contig of 708 bp in length  
\* 11718: 11817: gap of 100 bp  
\* 11818: 12474: contig of 657 bp in length  
\* 12475: 12574: gap of 100 bp  
\* 12575: 13236: contig of 662 bp in length  
\* 13237: 13336: gap of 100 bp  
\* 13337: 14011: contig of 675 bp in length  
\* 14011: 14111: gap of 100 bp  
\* 14112: 14820: contig of 709 bp in length  
\* 14821: 14920: gap of 100 bp  
\* 14921: 15634: contig of 714 bp in length  
\* 15635: 15734: gap of 100 bp

15735        \* 16412: contig of 678 bp in length  
             16513     gap of 100 bp  
             17218: contig of 706 bp in length  
             17318: gap of 100 bp  
             18010: contig of 692 bp in length  
             18110: gap of 100 bp  
             18790: contig of 680 bp in length  
             18890: gap of 100 bp  
             18891: contig of 704 bp in length  
             19595     19694: gap of 100 bp  
             20382: contig of 688 bp in length  
             20482: gap of 100 bp  
             21165: contig of 683 bp in length  
             21265: gap of 100 bp  
             21972: contig of 707 bp in length  
             21973     22072: gap of 100 bp  
             22753: contig of 681 bp in length  
             22754     22853: gap of 100 bp  
             23527: contig of 674 bp in length  
             23528     23627: gap of 100 bp  
             23628     24334: contig of 707 bp in length  
             24335     24434: gap of 100 bp  
             25133     25232: contig of 698 bp in length  
             25233     25910: gap of 100 bp  
             25911     26010: contig of 678 bp in length  
             26011     26691: contig of 681 bp in length  
             26692     27469: gap of 100 bp  
             27470     27569: gap of 100 bp  
             27570     28256: contig of 687 bp in length  
             28357     28356: gap of 100 bp  
             29071     29072: contig of 715 bp in length  
             29171     29868: gap of 100 bp  
             29969     29968: contig of 697 bp in length  
             30720     30719: contig of 751 bp in length  
             30820     31527: contig of 100 bp  
             31528     31627: contig of 708 bp in length  
             31628     32310: contig of 683 bp in length  
             32410     33099: gap of 100 bp  
             33100     33199: contig of 689 bp in length  
             33200     33877: contig of 678 bp in length  
             33978     34666: contig of 100 bp  
             34667     34766: gap of 100 bp  
             34767     35460: contig of 694 bp in length  
             35461     35560: gap of 100 bp  
             35561     36221: contig of 711 bp in length  
             36221     36372: gap of 100 bp  
             36372     37045: contig of 674 bp in length  
             37046     37145: gap of 100 bp  
             37146     37835: contig of 690 bp in length  
             37836     37935: gap of 100 bp  
             37936     38609: contig of 674 bp in length  
             38610     38709: gap of 100 bp  
             38710     39386: contig of 677 bp in length  
             39486     39486: gap of 100 bp  
             39487     40166: contig of 680 bp in length  
             40167     40266: gap of 100 bp  
             40267     40937: contig of 671 bp in length  
             40938     41037: gap of 100 bp  
             41038     41717: contig of 680 bp in length  
             41718     41817: gap of 100 bp  
             41818     42511: contig of 694 bp in length.

FEATURES  
source

1. .organism="Mus musculus"  
  mol\_type="genomic DNA"  
  db\_xref="taxon:10090"  
  clone="RP23-152017"

Query Match     100.0%; Score 231; DB 2; Length 42511;  
  Best Local Similarity 100.0%; Pred. No. 1.e-65;  
  Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY     1 GTGGGGTGGGAGCTTAAGCACTGGACAGAGGATAACCTGGTCATCCTGTGGAAGG 60  
  Db     27042 GTGGGGTGGGAGCTTAAGCACTGGACAGAGGATAACCTGGTCATCCTGTGGAAGG 271.01  
  QY     61 CACCTCTGATTGCTGAAGCTTCAGTGTCTACTCCCTCCCCAGAAAAAGACACG 120  
  Db     27102 CACGCTCTGATTGCTGAAGCTTCAGTGTCTACTCCCTCCCCAGAAAAAGACACG 271.61  
  QY     121 GGAGCTGGCCAAAGACCTCTGGGTGATGAGCCTAAGGGATGGTTGGTAGGCCCT 180  
  Db     27162 GGAGCTGGCCAAAGACCTCTGGGTGATGAGCCTAAGGGATGGTTGGTAGGCCCT 272.21  
  QY     181 ATGCUTGCACACTGGGATAGACCTGGGATCACCTCACCTCACCCATGAGGCTTGCTT 231  
  Db     27222 ATGCUTGCACACTGGGATAGACCTCACCTCACCCATGAGGCTTGCTT 272.72

RESULT 5  
  AC101394  
  LOCUS     64042 bp     DNA     linear     HTG 23 -NOV-2001  
  DEFINITION Mus musculus clone RP23-119F14, LOW-PASS SEQUENCE SAMPLING.  
  ACCESSION AC101394  
  VERSION AC101394.1     GI:17060169  
  KEYWORDS HTG; HTGS\_PHASE0.  
  SOURCE Mus musculus (house mouse)  
  ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
  Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
  REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
  AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
  TITLE Mus musculus, clone RP23-119F14  
  JOURNAL Unpublished  
  PUBLICATIONS 2 (bases 1 to 64042)  
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Boguski,V., Boukhalter,B., Brown,A., Camarota,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cool,A., Cooke,P., DeAngelis,K., Dellar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Goyette,M., Graham,L., Grand-Pierre,N., Ginde,S., Gord,S., Haford,A., Horizon,L., Hulme,W., Iliev,I., Jones,C., Kamat,A., Karats,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehozky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKerrian,K., McPheevers,R., Melidrim,J., Meneus,L., Mihora,T., Mleriga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norma,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rose,C., Rogov,P., Roman,J., Roset,M., Roy,A., Santo,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tsafave,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainon,J., Zemek,L., Zimmer,A. and Zody,M.

JOURNAL Direct Submission  
  COMMENT All repeats were identified using RepeatMasker:  
  http://ftp.genome.washington.edu/RM/RepeatMasker.html  
  COMMENT Center: Whitehead Institute/MIT Center for Genome Research  
  Center code: WIBR  
  Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L16611  
 Center clone name: 119\_F14

\* NOTE: This record contains 79 individual contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

|         |         |                            |
|---------|---------|----------------------------|
| 1       | 692:    | contig of 692 bp in length |
| 693     | 792:    | gap of 100 bp              |
| 793     | 151:    | contig of 719 bp in length |
| 151:    | 1611:   | gap of 100 bp              |
| 1611:   | 2313:   | contig of 702 bp in length |
| 2313:   | 2414:   | gap of 100 bp              |
| 2414:   | 3108:   | contig of 695 bp in length |
| 3108:   | 3208:   | gap of 100 bp              |
| 3208:   | 3920:   | contig of 721 bp in length |
| 3920:   | 4028:   | gap of 100 bp              |
| 4028:   | 4745:   | contig of 716 bp in length |
| 4745:   | 4845:   | gap of 100 bp              |
| 4845:   | 4846:   | contig of 714 bp in length |
| 4846:   | 5555:   | gap of 100 bp              |
| 5555:   | 5659:   | contig of 745 bp in length |
| 5659:   | 6404:   | gap of 100 bp              |
| 6404:   | 6505:   | contig of 731 bp in length |
| 6505:   | 7235:   | gap of 100 bp              |
| 7235:   | 7236:   | contig of 733 bp in length |
| 7236:   | 8066:   | gap of 100 bp              |
| 8066:   | 8168:   | contig of 707 bp in length |
| 8168:   | 8875:   | gap of 100 bp              |
| 8875:   | 8876:   | contig of 731 bp in length |
| 8876:   | 9697:   | gap of 100 bp              |
| 9697:   | 9797:   | contig of 711 bp in length |
| 9797:   | 10508:  | gap of 100 bp              |
| 10508:  | 10609:  | contig of 705 bp in length |
| 10609:  | 11313:  | gap of 100 bp              |
| 11313:  | 11314:  | contig of 722 bp in length |
| 11314:  | 8916:   | gap of 100 bp              |
| 8916:   | 9698:   | contig of 711 bp in length |
| 9698:   | 12211:  | gap of 100 bp              |
| 12211:  | 12212:  | contig of 715 bp in length |
| 12212:  | 13027:  | gap of 100 bp              |
| 13027:  | 13733:  | contig of 707 bp in length |
| 13733:  | 13734:  | gap of 100 bp              |
| 13734:  | 13834:  | contig of 711 bp in length |
| 13834:  | 14544:  | gap of 100 bp              |
| 14544:  | 14644:  | contig of 702 bp in length |
| 14644:  | 16970:  | gap of 100 bp              |
| 16970:  | 17070:  | contig of 701 bp in length |
| 17070:  | 17764:  | gap of 100 bp              |
| 17764:  | 18570:  | contig of 723 bp in length |
| 18570:  | 16268:  | gap of 100 bp              |
| 16268:  | 16269:  | contig of 719 bp in length |
| 16269:  | 18671:  | gap of 100 bp              |
| 18671:  | 19380:  | contig of 729 bp in length |
| 19380:  | 19499:  | gap of 100 bp              |
| 19499:  | 17011:  | contig of 694 bp in length |
| 17011:  | 20198:  | gap of 100 bp              |
| 20198:  | 21020:  | contig of 733 bp in length |
| 21020:  | 21131:  | gap of 100 bp              |
| 21131:  | 21855:  | contig of 729 bp in length |
| 21855:  | 21860:  | gap of 100 bp              |
| 21860:  | 22675:  | contig of 716 bp in length |
| 22675:  | 22776:  | gap of 100 bp              |
| 22776:  | 23497:  | contig of 721 bp in length |
| 23497:  | 23596:  | gap of 100 bp              |
| 23596:  | 24308:  | contig of 711 bp in length |
| 24308:  | 24408:  | gap of 100 bp              |
| 24408:  | 25127:  | contig of 720 bp in length |
| 25127:  | 25228:  | gap of 100 bp              |
| 25228:  | 25942:  | contig of 714 bp in length |
| 25942:  | 26042:  | gap of 100 bp              |
| 26042:  | 26758:  | contig of 717 bp in length |
| 26758:  | 26759:  | gap of 100 bp              |
| 26759:  | 26859:  | contig of 725 bp in length |
| 26859:  | 27584:  | gap of 100 bp              |
| 27584:  | 27684:  | contig of 712 bp in length |
| 27684:  | 28495:  | gap of 100 bp              |
| 28495:  | 29167:  | contig of 672 bp in length |
| 29167:  | 29168:  | gap of 100 bp              |
| 29168:  | 29268:  | contig of 651 bp in length |
| 29268:  | 29918:  | gap of 100 bp              |
| 29918:  | 30019:  | contig of 716 bp in length |
| 30019:  | 30735:  | gap of 100 bp              |
| 30735:  | 30835:  | contig of 738 bp in length |
| 30835:  | 31573:  | gap of 100 bp              |
| 31573:  | 31673:  | contig of 705 bp in length |
| 31673:  | 32378:  | gap of 100 bp              |
| 32378:  | 32477:  | contig of 716 bp in length |
| 32477:  | 33193:  | gap of 100 bp              |
| 33193:  | 33293:  | contig of 689 bp in length |
| 33293:  | 333294: | gap of 100 bp              |
| 333294: | 33982:  | contig of 689 bp in length |
| 33982:  | 34082:  | gap of 100 bp              |
| 34082:  | 34495:  | contig of 713 bp in length |
| 34495:  | 34795:  | gap of 100 bp              |
| 34795:  | 34895:  | contig of 708 bp in length |
| 34895:  | 35603:  | gap of 100 bp              |
| 35603:  | 35703:  | contig of 724 bp in length |
| 35703:  | 35704:  | gap of 100 bp              |
| 35704:  | 36427:  | contig of 721 bp in length |
| 36427:  | 36528:  | gap of 100 bp              |
| 36528:  | 37257:  | contig of 729 bp in length |
| 37257:  | 37357:  | gap of 100 bp              |
| 37357:  | 38079:  | contig of 723 bp in length |
| 38079:  | 38080:  | gap of 100 bp              |
| 38080:  | 38179:  | contig of 721 bp in length |
| 38179:  | 38900:  | gap of 100 bp              |
| 38900:  | 39001:  | contig of 717 bp in length |
| 39001:  | 39717:  | gap of 100 bp              |
| 39717:  | 39817:  | contig of 715 bp in length |
| 39817:  | 40532:  | gap of 100 bp              |
| 40532:  | 40533:  | contig of 704 bp in length |
| 40533:  | 40632:  | gap of 100 bp              |
| 40632:  | 40633:  | contig of 710 bp in length |
| 40633:  | 41343:  | gap of 100 bp              |
| 41343:  | 41442:  | contig of 696 bp in length |
| 41442:  | 42138:  | gap of 100 bp              |
| 42138:  | 42238:  | contig of 718 bp in length |
| 42238:  | 42942:  | gap of 100 bp              |
| 42942:  | 43042:  | contig of 699 bp in length |
| 43042:  | 43043:  | gap of 100 bp              |
| 43043:  | 43742:  | contig of 713 bp in length |
| 43742:  | 43841:  | gap of 100 bp              |
| 43841:  | 44559:  | contig of 715 bp in length |
| 44559:  | 44656:  | gap of 100 bp              |
| 44656:  | 45353:  | contig of 694 bp in length |
| 45353:  | 45354:  | gap of 100 bp              |
| 45354:  | 45454:  | contig of 721 bp in length |
| 45454:  | 46175:  | gap of 100 bp              |
| 46175:  | 46274:  | contig of 710 bp in length |
| 46274:  | 46675:  | gap of 100 bp              |
| 46675:  | 46989:  | contig of 707 bp in length |
| 46989:  | 47083:  | gap of 100 bp              |
| 47083:  | 47789:  | contig of 700 bp in length |
| 47789:  | 47790:  | gap of 100 bp              |
| 47790:  | 47890:  | contig of 713 bp in length |
| 47890:  | 48603:  | gap of 100 bp              |
| 48603:  | 48702:  | contig of 717 bp in length |
| 48702:  | 48703:  | gap of 100 bp              |
| 48703:  | 49409:  | contig of 720 bp in length |
| 49409:  | 49509:  | gap of 100 bp              |
| 49509:  | 49510:  | contig of 679 bp in length |
| 49510:  | 50188:  | contig of 704 bp in length |
| 50188:  | 50288:  | gap of 100 bp              |
| 50288:  | 51005:  | contig of 717 bp in length |
| 51005:  | 51105:  | gap of 100 bp              |
| 51105:  | 51106:  | contig of 722 bp in length |
| 51106:  | 51107:  | gap of 100 bp              |
| 51107:  | 51108:  | contig of 723 bp in length |
| 51108:  | 51109:  | gap of 100 bp              |
| 51109:  | 51110:  | contig of 724 bp in length |
| 51110:  | 51111:  | gap of 100 bp              |
| 51111:  | 51112:  | contig of 725 bp in length |
| 51112:  | 51113:  | gap of 100 bp              |
| 51113:  | 51114:  | contig of 726 bp in length |
| 51114:  | 51115:  | gap of 100 bp              |
| 51115:  | 51116:  | contig of 727 bp in length |
| 51116:  | 51117:  | gap of 100 bp              |
| 51117:  | 51118:  | contig of 728 bp in length |
| 51118:  | 51119:  | gap of 100 bp              |
| 51119:  | 51120:  | contig of 729 bp in length |
| 51120:  | 51121:  | gap of 100 bp              |
| 51121:  | 51122:  | contig of 730 bp in length |
| 51122:  | 51123:  | gap of 100 bp              |
| 51123:  | 51124:  | contig of 731 bp in length |
| 51124:  | 51125:  | gap of 100 bp              |
| 51125:  | 51126:  | contig of 732 bp in length |
| 51126:  | 51127:  | gap of 100 bp              |
| 51127:  | 51128:  | contig of 733 bp in length |
| 51128:  | 51129:  | gap of 100 bp              |
| 51129:  | 51130:  | contig of 734 bp in length |
| 51130:  | 51131:  | gap of 100 bp              |
| 51131:  | 51132:  | contig of 735 bp in length |
| 51132:  | 51133:  | gap of 100 bp              |
| 51133:  | 51134:  | contig of 736 bp in length |
| 51134:  | 51135:  | gap of 100 bp              |
| 51135:  | 51136:  | contig of 737 bp in length |
| 51136:  | 51137:  | gap of 100 bp              |
| 51137:  | 51138:  | contig of 738 bp in length |
| 51138:  | 51139:  | gap of 100 bp              |
| 51139:  | 51140:  | contig of 739 bp in length |
| 51140:  | 51141:  | gap of 100 bp              |
| 51141:  | 51142:  | contig of 740 bp in length |
| 51142:  | 51143:  | gap of 100 bp              |
| 51143:  | 51144:  | contig of 741 bp in length |
| 51144:  | 51145:  | gap of 100 bp              |
| 51145:  | 51146:  | contig of 742 bp in length |
| 51146:  | 51147:  | gap of 100 bp              |
| 51147:  | 51148:  | contig of 743 bp in length |
| 51148:  | 51149:  | gap of 100 bp              |
| 51149:  | 51150:  | contig of 744 bp in length |
| 51150:  | 51151:  | gap of 100 bp              |
| 51151:  | 51152:  | contig of 745 bp in length |
| 51152:  | 51153:  | gap of 100 bp              |
| 51153:  | 51154:  | contig of 746 bp in length |
| 51154:  | 51155:  | gap of 100 bp              |
| 51155:  | 51156:  | contig of 747 bp in length |
| 51156:  | 51157:  | gap of 100 bp              |
| 51157:  | 51158:  | contig of 748 bp in length |
| 51158:  | 51159:  | gap of 100 bp              |
| 51159:  | 51160:  | contig of 749 bp in length |
| 51160:  | 51161:  | gap of 100 bp              |
| 51161:  | 51162:  | contig of 750 bp in length |
| 51162:  | 51163:  | gap of 100 bp              |
| 51163:  | 51164:  | contig of 751 bp in length |
| 51164:  | 51165:  | gap of 100 bp              |
| 51165:  | 51166:  | contig of 752 bp in length |
| 51166:  | 51167:  | gap of 100 bp              |
| 51167:  | 51168:  | contig of 753 bp in length |
| 51168:  | 51169:  | gap of 100 bp              |
| 51169:  | 51170:  | contig of 754 bp in length |
| 51170:  | 51171:  | gap of 100 bp              |
| 51171:  | 51172:  | contig of 755 bp in length |
| 51172:  | 51173:  | gap of 100 bp              |
| 51173:  | 51174:  | contig of 756 bp in length |
| 51174:  | 51175:  | gap of 100 bp              |
| 51175:  | 51176:  | contig of 757 bp in length |
| 51176:  | 51177:  | gap of 100 bp              |
| 51177:  | 51178:  | contig of 758 bp in length |
| 51178:  | 51179:  | gap of 100 bp              |
| 51179:  | 51180:  | contig of 759 bp in length |
| 51180:  | 51181:  | gap of 100 bp              |
| 51181:  | 51182:  | contig of 760 bp in length |
| 51182:  | 51183:  | gap of 100 bp              |
| 51183:  | 51184:  | contig of 761 bp in length |
| 51184:  | 51185:  | gap of 100 bp              |
| 51185:  | 51186:  | contig of 762 bp in length |
| 51186:  | 51187:  | gap of 100 bp              |
| 51187:  | 51188:  | contig of 763 bp in length |
| 51188:  | 51189:  | gap of 100 bp              |
| 51189:  | 51190:  | contig of 764 bp in length |
| 51190:  | 51191:  | gap of 100 bp              |
| 51191:  | 51192:  | contig of 765 bp in length |
| 51192:  | 51193:  | gap of 100 bp              |
| 51193:  | 51194:  | contig of 766 bp in length |
| 51194:  | 51195:  | gap of 100 bp              |
| 51195:  | 51196:  | contig of 767 bp in length |
| 51196:  | 51197:  | gap of 100 bp              |
| 51197:  | 51198:  | contig of 768 bp in length |
| 51198:  | 51199:  | gap of 100 bp              |
| 51199:  | 51200:  | contig of 769 bp in length |
| 51200:  | 51201:  | gap of 100 bp              |
| 51201:  | 51202:  | contig of 770 bp in length |
| 51202:  | 51203:  | gap of 100 bp              |
| 51203:  | 51204:  | contig of 771 bp in length |
| 51204:  | 51205:  | gap of 100 bp              |
| 51205:  | 51206:  | contig of 772 bp in length |
| 51206:  | 51207:  | gap of 100 bp              |
| 51207:  | 51208:  | contig of 773 bp in length |
| 51208:  | 51209:  | gap of 100 bp              |
| 51209:  | 51210:  | contig of 774 bp in length |
| 51210:  | 51211:  | gap of 100 bp              |
| 51211:  | 51212:  | contig of 775 bp in length |
| 51212:  | 51213:  | gap of 100 bp              |
| 51213:  | 51214:  | contig of 776 bp in length |
| 51214:  | 51215:  | gap of 100 bp              |
| 51215:  | 51216:  | contig of 777 bp in length |
| 51216:  | 51217:  | gap of 100 bp              |
| 51217:  | 51218:  | contig of 778 bp in length |
| 51218:  | 51219:  | gap of 100 bp              |
| 51219:  | 51220:  | contig of 779 bp in length |
| 51220:  | 51221:  | gap of 100 bp              |
| 51221:  | 51222:  | contig of 780 bp in length |
| 51222:  | 51223:  | gap of 100 bp              |
| 51223:  | 51224:  | contig of 781 bp in length |
| 51224:  | 51225:  | gap of 100 bp              |
| 51225:  | 51226:  | contig of 782 bp in length |
| 51226:  | 51227:  | gap of 100 bp              |
| 51227:  | 51228:  | contig of 783 bp in length |
| 51228:  | 51229:  | gap of 100 bp              |
| 51229:  | 51230:  | contig of 784 bp in length |
| 51230:  | 51231:  | gap of 100 bp              |
| 51231:  | 51232:  | contig of 785 bp in length |
| 51232:  | 51233:  | gap of 100 bp              |
| 51233:  | 51234:  | contig of 786 bp in length |
| 51234:  | 51235:  | gap of 100 bp              |
| 51235:  | 51236:  | contig of 787 bp in length |
| 51236:  | 51237:  | gap of 100 bp              |
| 51237:  | 51238:  | contig of 788 bp in length |
| 51238:  | 51239:  | gap of 100 bp              |
| 51239:  | 51240:  | contig of 789 bp in length |
| 51240:  | 51241:  | gap of 100 bp              |
| 51241:  | 51242:  | contig of 790 bp in length |
| 51242:  | 51243:  | gap of 100 bp              |
| 51243:  | 51244:  | contig of 791 bp in length |
| 51244:  | 51245:  | gap of 100 bp              |
| 51245:  | 51246:  | contig of 792 bp in length |
| 51246:  | 51247:  | gap of 100 bp              |
| 51247:  | 51248:  | contig of 793 bp in length |
| 51248:  | 51249:  | gap of 100 bp              |
| 51249:  | 51250:  | contig of 794 bp in length |
| 51250:  | 51251:  | gap of 100 bp              |
| 51251:  | 51252:  | contig of 795 bp in length |
| 51252:  | 51253:  | gap of 100 bp              |
| 51253:  | 51254:  | contig of 796 bp in length |
| 51254:  | 51255:  | gap of 100 bp              |
| 51255:  | 51256:  | contig of 797 bp in length |
| 51256:  | 51257:  | gap of 100 bp              |
| 51257:  | 51258:  | contig of 798 bp in length |
| 51258:  | 51259:  | gap of 100 bp              |
| 51259:  | 51260:  | contig of 799 bp in length |
| 51260:  | 51261:  | gap of 100 bp              |
| 51261   |         |                            |

Query Match 100.0%; Score 231; DB 2; Length 64042;

Best Local Similarity 100.0%; Pred No. 1.7e-65;

Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAACGACTGACAGAGGATAAGCTGCTGTTGGCATCCCTGGAAAGG 60  
 DB 32664 GTGGGTGCGAGGCTAACGACTGACAGAGGATAAGCTGCTGTTGGCATCCCTGGAAAGG 32723

QY 61 CAGCTCTATTGATGATGAGGTTCAACGACTCACAGAGGATAAGCTGCTGTTGGCATCCCTGGAAAGG 120  
 DB 32724 CAGCTCTATTGATGATGAGGTTCAACGACTCACAGAGGATAAGCTGCTGTTGGCATCCCTGGAAAGG 32783

QY 121 GGAGCTGGCAAGACCTCTGGTGAAGCTTAAGGATGGTTTGTAAGGCCCT 180  
 DB 32784 GGAGCTGGCAAGACCTCTGGTGAAGCTTAAGGATGGTTTGTAAGGCCCT 32843

QY 181 ATGCTTGACACNGGGATCAGACCTCACCCATGAGGCTAGTCCCTCCAGAAAGCAGACG 231  
 DB -32844 ATGCTTGACACNGGGATCAGCCATGAGGCTAGTCCCTCCAGAAAGCAGACG 32894

---

RESULT 6

AC100424/c LOCUS Mus musculus clone RP23-136F8, DNA linear HTG 22-NOV-2001

DEFINITION Mus musculus clone RP23-136F8, LOW-PASS SEQUENCE SAMPLING.

AC100424 VERSION AC100424.1 GI:17047790

HTG; HTGS; PHASED

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 66393)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Mus musculus, clone RP23-136F8

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 66393)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Landier,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gade,D., Galatyn,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Haggos,B., Heaford,A., Horton,I., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,J., LaRocque,K., Lamazares,R., Landers,T., Lehozky,J., Levine,R., Liu,G., Maclean,C., McDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheebers,R., Meldrum,J., Menecus,L., Milova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Reitta,R., Riley,R., Rileback,M., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schupack,R., Seaman,S., Severy,P., Subramanian,A., Talama,J., Testayre,S., Theodore,J., Strauss,N., Stange-Thomann,N., Stojanovic,N., Spangler,B., Talamas,J., Trigilio,J., Vassiliev,H., Topham,K., Travers,M., Travis,N., Wyman,D., Ye,W.J., Young,G., Vieil,R., Vo,A., Wilcox,J., Wu,X., Zembek,L., Zimmer,A. and Zody,M.

JOURNAL Direct Submission

COMMENT Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L15312  
 Center clone name: 136\_F\_8

\* NOTE: This record contains 84 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequencing is useful for identifying clones that may be gene-rich and allows overal relationships among clones to be deduced.  
 However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 692: contig of 692 bp in length  
 \* 693: gap of 100 bp  
 \* 693: 1509: contig of 717 bp in length  
 \* 1510: 1609: gap of 100 bp  
 \* 1610: 2315: contig of 706 bp in length  
 \* 2316: 2415: gap of 100 bp  
 \* 2416: 3132: contig of 717 bp in length  
 \* 3133: 3232: gap of 100 bp  
 \* 3233: 3912: contig of 680 bp in length  
 \* 3913: 4012: gap of 100 bp  
 \* 4013: 4682: contig of 670 bp in length  
 \* 4683: 4782: gap of 100 bp  
 \* 4783: 5477: contig of 695 bp in length  
 \* 5478: 5578: 6280: contig of 703 bp in length  
 \* 6281: 6380: gap of 100 bp  
 \* 6381: 7066: contig of 686 bp in length  
 \* 7067: 7166: gap of 100 bp  
 \* 7167: 7870: contig of 704 bp in length  
 \* 7871: 8970: gap of 100 bp  
 \* 8666: 8765: contig of 695 bp in length  
 \* 8766: 9472: contig of 707 bp in length  
 \* 9473: 9573: 10230: contig of 658 bp in length  
 \* 10231: 10330: gap of 100 bp  
 \* 10331: 11024: contig of 694 bp in length  
 \* 11124: 11125: 11809: contig of 685 bp in length  
 \* 11810: 11909: gap of 100 bp  
 \* 12589: 12689: 13269: contig of 680 bp in length  
 \* 12689: 13274: contig of 685 bp in length  
 \* 13274: 13474: gap of 100 bp  
 \* 13475: 14170: contig of 696 bp in length  
 \* 14171: 14270: gap of 100 bp  
 \* 14271: 14967: contig of 697 bp in length  
 \* 14968: 15067: gap of 100 bp  
 \* 15068: 15770: contig of 703 bp in length  
 \* 15771: 15870: gap of 100 bp  
 \* 15871: 16588: contig of 718 bp in length  
 \* 16589: 16688: gap of 100 bp  
 \* 16689: 17357: contig of 669 bp in length  
 \* 17358: 17457: gap of 100 bp  
 \* 17458: 18160: contig of 703 bp in length  
 \* 18161: 18260: gap of 100 bp  
 \* 18261: 18972: contig of 712 bp in length  
 \* 18973: 19072: gap of 100 bp  
 \* 19073: 19724: contig of 652 bp in length  
 \* 19725: 19824: gap of 100 bp  
 \* 19825: 20520: contig of 696 bp in length  
 \* 20521: 20620: gap of 100 bp

|       |                                   |   |       |                                   |  |
|-------|-----------------------------------|---|-------|-----------------------------------|--|
| 20621 | 21326: contig of 706 bp in length | * | 49775 | 49874: gap of 100 bp              |  |
| 21327 | 21426: gap of 100 bp              | * | 49875 | 50562: contig of 688 bp in length |  |
| 21427 | 22117: contig of 691 bp in length | * | 50563 | 50662: gap of 100 bp              |  |
| 22118 | 22218: gap of 100 bp              | * | 50663 | 51336: contig of 674 bp in length |  |
| 22218 | 22904: contig of 687 bp in length | * | 51337 | 51436: gap of 100 bp              |  |
| 22905 | 23004: gap of 100 bp              | * | 51437 | 52136: contig of 700 bp in length |  |
| 23005 | 23705: contig of 701 bp in length | * | 52137 | 52236: gap of 100 bp              |  |
| 23705 | 23806: gap of 100 bp              | * | 52237 | 52967: contig of 731 bp in length |  |
| 23806 | 24497: contig of 695 bp in length | * | 52968 | 53067: gap of 100 bp              |  |
| 24498 | 24597: gap of 100 bp              | * | 53068 | 53148: contig of 681 bp in length |  |
| 24598 | 25272: contig of 675 bp in length | * | 53749 | 53848: gap of 100 bp              |  |
| 25273 | 25372: gap of 100 bp              | * | 53849 | 54539: contig of 691 bp in length |  |
| 25373 | 26083: contig of 711 bp in length | * | 54540 | 54639: gap of 100 bp              |  |
| 26084 | 26183: gap of 100 bp              | * |       |                                   | Query Match Score 100.0%; Pred. No. 1  |
| 26184 | 26860: contig of 677 bp in length | * |       |                                   | e-65; DB 2; Length 66393;  |
| 26861 | 26960: gap of 100 bp              | * |       |                                   | Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;  |
| 26961 | 27628: contig of 668 bp in length | * |       |                                   | Matches 231; Conservative 0;   |
| 27629 | 27728: gap of 100 bp              | * |       |                                   | AC101278 AC101278  |
| 27729 | 28417: contig of 689 bp in length | * |       |                                   | Mus musculus clone RP23-101P16, LOW PASS SEQUENCE SAMPLING.  |
| 28418 | 28517: gap of 100 bp              | * |       |                                   | AC101278.1 GI:17060053   |
| 28518 | 29199: contig of 682 bp in length | * |       |                                   | HTG; HTGS; PHASE0.   |
| 29200 | 29299: gap of 100 bp              | * |       |                                   | Mus musculus (house mouse)   |
| 29957 | 30057: contig of 658 bp in length | * |       |                                   | Unpublished  |
| 29958 | 30057: gap of 100 bp              | * |       |                                   | Organism   |
| 30058 | 30742: contig of 685 bp in length | * |       |                                   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |
| 30743 | 30842: gap of 100 bp              | * |       |                                   | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |
| 30843 | 31537: contig of 695 bp in length | * |       |                                   | REFERENCE 1 (bases 1 to 66489)   |
| 31538 | 31637: gap of 100 bp              | * |       |                                   | AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.   |
| 31638 | 32348: contig of 711 bp in length | * |       |                                   | JOURNAL  |
| 32349 | 32448: gap of 100 bp              | * |       |                                   | RESULT 7   |
| 32449 | 33128: contig of 680 bp in length | * |       |                                   | AC101278   |
| 33129 | 33228: gap of 100 bp              | * |       |                                   | LOCUS AC101278   |
| 33229 | 33910: contig of 682 bp in length | * |       |                                   | DEFINITION Mus musculus  |
| 33910 | 34010: gap of 100 bp              | * |       |                                   | VERSION AC101278   |
| 34011 | 34723: contig of 713 bp in length | * |       |                                   | ACCESSION AC101278.1   |
| 34724 | 34823: gap of 100 bp              | * |       |                                   | KEYWORDS HTG; HTGS; PHASE0.  |
| 34824 | 35524: contig of 701 bp in length | * |       |                                   | SOURCE Mus musculus  |
| 35525 | 35624: gap of 100 bp              | * |       |                                   | ORGANISM Mus musculus  |
| 35625 | 36304: contig of 680 bp in length | * |       |                                   | REFERENCE 2 (bases 1 to 66489)   |
| 36305 | 36404: gap of 100 bp              | * |       |                                   | AUTHORS Anderson, S., Barton, N., Bastien, V., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chaar, B., Choepel, Y., Coangello, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferraira, P., FitzHugh, W., Gage, D., Galagan, J., Garyns, S., Ginde, S., Gord, S., Goyet, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Huime, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karats, A., Keilis, C., Larocque, K., Lamazares, R., Landers, T., Lehozcy, J., Levine, R., Liu, G., MacLean, C., McDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeeters, R., Medrim, J., Menes, L., Mihova, T., Mielinga, V., Murphy, J., Mayor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, R., Riley, R., Roy, A., Santos, R., Schaefer, S., Schupback, R., Roman, J., Roselli, M., Seaman, S., Seaver, P., Spander, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. |
| 39435 | 39534: gap of 100 bp              | * |       |                                   | REFERENCE 3 (bases 1 to 66489)   |
| 39535 | 40251: contig of 717 bp in length | * |       |                                   | AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.   |
| 40252 | 40351: gap of 100 bp              | * |       |                                   | JOURNAL  |
| 40352 | 41047: contig of 696 bp in length | * |       |                                   | RESULT 7   |
| 41048 | 41147: gap of 100 bp              | * |       |                                   | AC101278   |
| 41148 | 41821: contig of 674 bp in length | * |       |                                   | LOCUS AC101278   |
| 41822 | 41921: gap of 100 bp              | * |       |                                   | DEFINITION Mus musculus  |
| 41922 | 42631: contig of 710 bp in length | * |       |                                   | VERSION AC101278.1   |
| 42632 | 42731: gap of 100 bp              | * |       |                                   | ACCESSION AC101278.1   |
| 42732 | 43437: contig of 706 bp in length | * |       |                                   | KEYWORDS HTG; HTGS; PHASE0.  |
| 43438 | 4353: gap of 100 bp               | * |       |                                   | SOURCE Mus musculus  |
| 43538 | 44243: contig of 706 bp in length | * |       |                                   | ORGANISM Mus musculus  |
| 44244 | 44343: gap of 100 bp              | * |       |                                   | REFERENCE 2 (bases 1 to 66489)   |
| 44344 | 45011: contig of 668 bp in length | * |       |                                   | AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.   |
| 45012 | 45112: gap of 100 bp              | * |       |                                   | JOURNAL  |
| 45112 | 45789: contig of 678 bp in length | * |       |                                   | RESULT 7   |
| 45790 | 4589: gap of 100 bp               | * |       |                                   | AC101278   |
| 45890 | 46583: contig of 694 bp in length | * |       |                                   | LOCUS AC101278   |
| 46584 | 46683: gap of 100 bp              | * |       |                                   | DEFINITION Mus musculus  |
| 46684 | 47363: contig of 680 bp in length | * |       |                                   | VERSION AC101278.1   |
| 47364 | 47463: gap of 100 bp              | * |       |                                   | ACCESSION AC101278.1   |
| 47464 | 48166: contig of 703 bp in length | * |       |                                   | KEYWORDS HTG; HTGS; PHASE0.  |
| 48167 | 48266: gap of 100 bp              | * |       |                                   | SOURCE Mus musculus  |
| 48267 | 48971: contig of 705 bp in length | * |       |                                   | ORGANISM Mus musculus  |
| 48972 | 49071: gap of 100 bp              | * |       |                                   | REFERENCE 2 (bases 1 to 66489)   |
| 49074 | 49774: contig of 703 bp in length | * |       |                                   | AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.   |

**JOURNAL** Direct Submission (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

---

**Center:** Whitehead Institute/ MIT Center for Genome Research  
**Center code:** WIBR  
**Web site:** <http://www-seq.wi.mit.edu>  
**Contact:** sequence\_submissions@genome.wi.mit.edu  
**Project Information:**  
 Center Project name: L16354  
 Center clone name: 101\_P\_16

---

\* NOTE: This record contains 81 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 723: contig of 723 bp in length  
 \* 724 823: gap of 100 bp  
 \* 1542: contig of 719 bp in length  
 \* 1642: gap of 100 bp  
 \* 2341: contig of 699 bp in length  
 \* 2441: gap of 100 bp  
 \* 3174: contig of 733 bp in length  
 \* 3274: gap of 100 bp  
 \* 3275 4014: contig of 740 bp in length  
 \* 4114: gap of 100 bp  
 \* 4846: contig of 732 bp in length  
 \* 4847 4946: gap of 100 bp  
 \* 4947 5656: contig of 710 bp in length  
 \* 5657 5756: gap of 100 bp  
 \* 6473: contig of 717 bp in length  
 \* 6573: gap of 100 bp  
 \* 6574 7317: contig of 744 bp in length  
 \* 7318 7417: gap of 100 bp  
 \* 7418 8129: contig of 712 bp in length  
 \* 8130 8229: gap of 100 bp  
 \* 8230 8953: contig of 724 bp in length  
 \* 8954 9053: gap of 100 bp  
 \* 9054 9838: contig of 785 bp in length  
 \* 9839 9938: gap of 100 bp  
 \* 9939 10667: contig of 729 bp in length  
 \* 10668 10767: gap of 100 bp  
 \* 10768 11503: contig of 736 bp in length  
 \* 11504 11603: gap of 100 bp  
 \* 11604 12346: contig of 743 bp in length  
 \* 12347 12446: gap of 100 bp  
 \* 12447 13168: contig of 722 bp in length  
 \* 13169 13268: gap of 100 bp  
 \* 13269 13995: contig of 727 bp in length  
 \* 13996 14095: gap of 100 bp  
 \* 14096 14810: contig of 715 bp in length  
 \* 14811 14910: gap of 100 bp  
 \* 14911 15587: contig of 677 bp in length  
 \* 15588 15687: gap of 100 bp  
 \* 15688 16390: contig of 703 bp in length  
 \* 16391 16490: gap of 100 bp  
 \* 16491 17212: contig of 722 bp in length  
 \* 17213 17312: gap of 100 bp  
 \* 17313 18033: contig of 721 bp in length  
 \* 18034 18133: gap of 100 bp  
 \* 18134 18856: contig of 723 bp in length  
 \* 18857 18956: gap of 100 bp  
 \* 18957 19693: contig of 737 bp in length

19694 19793: gap of 100 bp  
 \* 19794 20529: contig of 736 bp in length  
 \* 20530 20630: gap of 100 bp  
 \* 21364: contig of 735 bp in length  
 \* 21365 21464: gap of 100 bp  
 \* 22465 22289: contig of 725 bp in length  
 \* 22290 22981: contig of 100 bp  
 \* 22982 23081: gap of 100 bp  
 \* 23082 23780: contig of 699 bp in length  
 \* 23781 23880: gap of 100 bp  
 \* 23881 24606: contig of 726 bp in length  
 \* 24607 24707: gap of 100 bp  
 \* 24707 25424: contig of 718 bp in length  
 \* 25425 25524: gap of 100 bp  
 \* 25525 26243: contig of 719 bp in length  
 \* 26244 26543: gap of 100 bp  
 \* 26544 27068: contig of 725 bp in length  
 \* 27069 27168: gap of 100 bp  
 \* 27169 27900: contig of 732 bp in length  
 \* 27901 28000: gap of 100 bp  
 \* 28001 28726: contig of 726 bp in length  
 \* 28727 28826: gap of 100 bp  
 \* 28827 29359: contig of 733 bp in length  
 \* 29359 29560: gap of 100 bp  
 \* 29560 30397: contig of 738 bp in length  
 \* 30398 30497: gap of 100 bp  
 \* 30498 31212: contig of 715 bp in length  
 \* 31213 31312: gap of 100 bp  
 \* 31313 32025: contig of 713 bp in length  
 \* 32026 32125: gap of 100 bp  
 \* 32126 32855: contig of 730 bp in length  
 \* 32855 32955: gap of 100 bp  
 \* 32956 33674: contig of 719 bp in length  
 \* 33675 33774: gap of 100 bp  
 \* 33775 34205: contig of 706 bp in length  
 \* 34205 34480: gap of 100 bp  
 \* 34481 34580: gap of 100 bp  
 \* 34581 35294: contig of 714 bp in length  
 \* 35295 35394: gap of 100 bp  
 \* 35395 36121: contig of 727 bp in length  
 \* 36122 36221: gap of 100 bp  
 \* 36222 36952: contig of 731 bp in length  
 \* 36952 37052: gap of 100 bp  
 \* 37053 37771: contig of 719 bp in length  
 \* 37772 37871: gap of 100 bp  
 \* 37872 38601: contig of 730 bp in length  
 \* 38602 38701: gap of 100 bp  
 \* 38702 39431: contig of 730 bp in length  
 \* 39432 39532: gap of 100 bp  
 \* 39532 40256: contig of 725 bp in length  
 \* 40257 40356: gap of 100 bp  
 \* 40357 41067: contig of 711 bp in length  
 \* 41068 41167: gap of 100 bp  
 \* 41168 41897: contig of 721 bp in length  
 \* 41897 41997: gap of 100 bp  
 \* 41998 42712: contig of 715 bp in length  
 \* 42713 42812: gap of 100 bp  
 \* 42812 43533: contig of 721 bp in length  
 \* 43533 43633: gap of 100 bp  
 \* 43634 44354: contig of 721 bp in length  
 \* 44355 44454: gap of 100 bp  
 \* 44455 45187: contig of 733 bp in length  
 \* 45188 45287: gap of 100 bp  
 \* 45287 46026: contig of 739 bp in length  
 \* 46027 46126: gap of 100 bp  
 \* 46127 46824: contig of 698 bp in length  
 \* 46825 46924: gap of 100 bp  
 \* 46925 47656: contig of 732 bp in length  
 \* 47657 47756: gap of 100 bp  
 \* 47757 48484: contig of 728 bp in length  
 \* 48484 48584: gap of 100 bp  
 \* 48585 49301: contig of 717 bp in length  
 \* 49301 49401: gap of 100 bp

Query Match      100.0%; Score: 231; DB: 2; Length: 66489;  
 Best Local Similarity: 100.0%; Pred. No. 1.7e-65;  
 Matches: 231; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

Qy      1 GTGGGTGCGAGGGCTAAGGACTGCACAGGATAGCTTGCTGTTGGCATCTGGAGG 60  
 Db      19918 GTGGGTGCGAGGGCTAAGGACTGCACAGGATAGCTTGCTGTTGGCATCTGGAGG 19977

Qy      61 CACGCTGATTGCGATGAAGTTCAAGTGTAGCTAGTCCTAGCTGCTTCAAGGAAACGACAG 120  
 Db      19978 CACGCTGATTGCGATGAAGTTCAAGTGTAGCTGCTTCAAGGAAACGACAG 20037

Qy      121 GGAGGTGCCAAGACCTCAGCTGGGTAGAGCCTAAGGGATGGTTGTAGGCCCT 180  
 Db      20038 GGAGTGGCCANGACTCTGGGTAGCTGGTGTAGGGTTGTAGGCCCT 20097

Qy      181 ATGCCTTGACACTGGGATCAGACCTCTACCCATGAGGCTTGCT 231  
 Db      20098 ATGCCTTGACACTGGGATCAGACCTCTACCCATGAGGCTTGCT 20148

RESULT 8  
 AC101278/C  
 DEFINITION Mus musculus clone RP23-101P16, LINEAR HTG 23-NOV-2001  
 AC101278 LOCUS Mus musculus clone RP23-101P16, LOW-PASS SEQUENCE SAMPLING.  
 AC101278.1 VERSION GI:17060053  
 HTG; HTGS\_PHASE0.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 66489)  
 Birren,B., Linton,L., Nusbaum,C., and Lander,E.  
 Unpublished  
 2 (bases 1 to 66489)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhvalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Huime,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Keils,C., LaRocque,K.,  
 Lamazares,R., Landers,T., Lehozczky,J., Levine,R., Liu,G.,  
 MacLean,C., McDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Mejdrim,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schupback,R.,  
 Schuster,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfae,S., Theodore,J.,  
 Topham,R., Travers,M., Travis,N., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Viel,R., Vo,A., Wilson,B., Ziemek,L., Zimmer,A., and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.wi.mit.edu/RW/RepeatMasker.html  
 CENTER CENTER CODE: WIBR  
 WEB SITE: http://www-seq.wi.mit.edu  
 CONTACT: sequence\_submissions@genome.wi.mit.edu  
 PROJECT INFORMATION  
 CENTER PROJECT NAME: L16354  
 CENTER CLONE NAME: 101\_P\_16

\* NOTE: This record contains 81 individual contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1      723: contig of 723 bp in length  
 \*      823: gap of 100 bp  
 \*      824: contig of 719 bp in length  
 \*      1543: gap of 100 bp  
 \*      1642: contig of 699 bp in length  
 \*      2341: contig of 699 bp in length  
 \*      2342: gap of 100 bp  
 \*      3174: contig of 733 bp in length  
 \*      3175: gap of 100 bp  
 \*      4014: contig of 740 bp in length  
 \*      4015: gap of 100 bp  
 \*      4115: contig of 732 bp in length  
 \*      4846: contig of 732 bp in length  
 \*      4947: gap of 100 bp  
 \*      5656: contig of 710 bp in length  
 \*      5657: gap of 100 bp  
 \*      5757: gap of 100 bp  
 \*      6473: gap of 100 bp  
 \*      6573: gap of 100 bp  
 \*      6574: contig of 744 bp in length  
 \*      7318: gap of 100 bp  
 \*      7417: gap of 100 bp  
 \*      7418: contig of 712 bp in length  
 \*      8130: gap of 100 bp  
 \*      8230: gap of 100 bp  
 \*      8953: contig of 724 bp in length  
 \*      9053: gap of 100 bp  
 \*      9054: contig of 785 bp in length  
 \*      9839: gap of 100 bp  
 \*      9939: contig of 729 bp in length  
 \*      10668: gap of 100 bp  
 \*      11503: contig of 736 bp in length  
 \*      11504: gap of 100 bp  
 \*      11604: contig of 743 bp in length  
 \*      12346: gap of 100 bp  
 \*      12347: contig of 722 bp in length  
 \*      13168: contig of 722 bp in length  
 \*      13169: gap of 100 bp  
 \*      13269: contig of 727 bp in length  
 \*      13995: gap of 100 bp  
 \*      14095: gap of 100 bp  
 \*      14096: contig of 715 bp in length  
 \*      14810: contig of 715 bp in length  
 \*      14811: gap of 100 bp  
 \*      14911: contig of 677 bp in length  
 \*      15587: contig of 677 bp in length  
 \*      15588: gap of 100 bp  
 \*      16390: contig of 703 bp in length  
 \*      16391: gap of 100 bp  
 \*      16490: gap of 100 bp  
 \*      16491: contig of 722 bp in length  
 \*      17212: contig of 722 bp in length  
 \*      17312: gap of 100 bp

17313: contig of 721 bp in length  
 18033: gap of 100 bp  
 18133: contig of 723 bp in length  
 18834: gap of 100 bp  
 18856: contig of 737 bp in length  
 18957: gap of 100 bp  
 19693: contig of 736 bp in length  
 19733: gap of 100 bp  
 19794: contig of 735 bp in length  
 20529: gap of 100 bp  
 20530: contig of 735 bp in length  
 21364: gap of 100 bp  
 21465: contig of 725 bp in length  
 22189: gap of 100 bp  
 22289: contig of 692 bp in length  
 22290: gap of 100 bp  
 23081: contig of 699 bp in length  
 23780: contig of 699 bp in length  
 23880: gap of 100 bp  
 23881: contig of 726 bp in length  
 24606: gap of 100 bp  
 24707: contig of 718 bp in length  
 25424: gap of 100 bp  
 25435: contig of 719 bp in length  
 25525: gap of 100 bp  
 26243: contig of 719 bp in length  
 26244: gap of 100 bp  
 26344: contig of 725 bp in length  
 27069: gap of 100 bp  
 27168: contig of 732 bp in length  
 27900: contig of 725 bp in length  
 27901: gap of 100 bp  
 28000: contig of 726 bp in length  
 28726: contig of 726 bp in length  
 28727: gap of 100 bp  
 28847: contig of 733 bp in length  
 29550: gap of 100 bp  
 29660: contig of 738 bp in length  
 30398: gap of 100 bp  
 30498: contig of 715 bp in length  
 31212: gap of 100 bp  
 31313: contig of 713 bp in length  
 32026: gap of 100 bp  
 32126: contig of 730 bp in length  
 32855: gap of 100 bp  
 32955: contig of 719 bp in length  
 33121: gap of 100 bp  
 33133: contig of 706 bp in length  
 33375: gap of 100 bp  
 34481: contig of 714 bp in length  
 34500: gap of 100 bp  
 34581: contig of 714 bp in length  
 35295: gap of 100 bp  
 35335: contig of 727 bp in length  
 36121: gap of 100 bp  
 36222: contig of 731 bp in length  
 36932: gap of 100 bp  
 37052: contig of 719 bp in length  
 37772: gap of 100 bp  
 37872: contig of 730 bp in length  
 38601: gap of 100 bp  
 38701: contig of 730 bp in length  
 38702: gap of 100 bp  
 38703: contig of 730 bp in length  
 39432: gap of 100 bp  
 39532: contig of 725 bp in length  
 40256: gap of 100 bp  
 40257: contig of 711 bp in length  
 40337: gap of 100 bp  
 41067: contig of 721 bp in length  
 41168: gap of 100 bp  
 41898: gap of 100 bp  
 41998: contig of 715 bp in length  
 42713: gap of 100 bp  
 42813: contig of 721 bp in length  
 43533: contig of 721 bp in length  
 43534: gap of 100 bp  
 44334: contig of 721 bp in length  
 44454: gap of 100 bp  
 44518: contig of 733 bp in length  
 45188: gap of 100 bp  
 45287: contig of 739 bp in length  
 46026: gap of 100 bp  
 46127: gap of 100 bp  
 46224: contig of 698 bp in length  
 46924: gap of 100 bp  
 46925: contig of 732 bp in length  
 46935: contig of 732 bp in length

47657: gap of 100 bp  
 47757: contig of 728 bp in length  
 48484: gap of 100 bp  
 48584: contig of 717 bp in length  
 48885: gap of 100 bp  
 49302: gap of 100 bp  
 49401: contig of 718 bp in length  
 49402: contig of 719 bp in length  
 50120: gap of 100 bp  
 50220: contig of 715 bp in length  
 50934: contig of 100 bp  
 50935: gap of 100 bp  
 51034: gap of 100 bp  
 51035: contig of 732 bp in length  
 51167: gap of 100 bp  
 51866: gap of 100 bp  
 52575: contig of 709 bp in length  
 52675: gap of 100 bp  
 53400: contig of 725 bp in length  
 53401: gap of 100 bp  
 53421: contig of 717 bp in length  
 54218: gap of 100 bp  
 54317: contig of 686 bp in length  
 54418: gap of 100 bp  
 55004: gap of 100 bp  
 55103: gap of 100 bp  
 55811: contig of 708 bp in length  
 55812: gap of 100 bp  
 55912: contig of 723 bp in length  
 56635: gap of 100 bp  
 56634: gap of 100 bp  
 56734: gap of 100 bp

Query Match Score 231; DB 2; Length 66489;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-65;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCCAGGGCTAACGACTGCCAGAGGATAGCTTGCTGGCATCCTGTGGAGG 60  
 Db 44342 GTGGGTGCCAGGTAAAGCACTGCCAGACTGCCAGAGGATAGCTTGCTGGCATCCTGTGGAGG 44283

Qy 61 CACGTGTGATGTCAGTGAAGGTTCACTGTGCCATAGTCCATTGCTTCCAGGAAAAACGACAGC 120  
 Db 44282 CACGTCTGATGTCAGTGAAGGTTCACTGTGCCATAGTGCCTTCCAGGAAAAACGACAGC 44223

Qy 121 GGAGCTGGCCAAGACCTCTGGGATCAGACCCCTAACCTTACCCATGAGGCTTAAGGGATGTTGGCTTAGGCCCT 180  
 Db 44222 GGAGCTGGCCAAGACCTCTGGGATCAGACCCCTAACCTTACCCATGAGGCTTAAGGGATGTTGGCTTAGGCCCT 44163

Qy 181 ATGCTTGCAACATGGGGATCAGACCCCTAACCTTACCCATGAGGCTTAAGGGATGTTGGCTTAGGCCCT 231  
 Db 44162 ATGCTTGCAACATGGGGATCAGACCCCTAACCTTACCCATGAGGCTTAAGGGATGTTGGCTTAGGCCCT 44112

RESULT 9 AC101122/c

LOCUS AC101122 70360 bp DNA linear HTG 23 -NOV-2001

DEFINITION Mus musculus clone RP23-93E18, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC101122

VERSION AC101122.1 GI:17059896

KEYWORDS HRG; HRCS\_PHASE0.

SOURCE Mus musculus (house mouse)

ORGANISM Mammalia; Eutheria; Rodentia; Craniata; Vertebrata; Euteleostomi; Eukaryota; Muridae; Murinae; Mus

REFERENCE 1 (bases 1 to 70360)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 70360)

AUTHORS Anderson, S., Barrera, N., Bastien, V., Boguslavsky, L., Boukgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAngelis, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hedges, B., Heaford, A., Horton, L., Huine, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoche, K., Lamazares, R., Launders, T., Lehozky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.N., Stojanovic, N., Strauss, N., Subramanian, A., Talanias, J., Testayre, S., Theodore, J., Topham, K., Travers, M., Travis, N., Tricilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** All repeats were identified using RepeatMasker:  
<http://fp.genome.washington.edu/RM/RepeatMasker.html>  
**Smit, A.F.A. & Green, P. (1996-1997)**  
**Center**: Whitehead Institute/ MIT Center for Genome Research  
**Center code**: WIBR  
**Web site**: <http://www-seq.wi.mit.edu>  
**Contact**: sequence\_submissions@genome.wi.mit.edu  
**Project Information**  
**Center project name**: L14388  
**Center clone name**: 93\_E\_18

---

\* NOTE: This record contains 86 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

\* 721: contig of 721 bp in length  
\* 821: gap of 100 bp  
\* 1528: contig of 707 bp in length  
\* 1529: gap of 100 bp  
\* 2332: contig of 704 bp in length  
\* 2333: gap of 100 bp  
\* 2432: contig of 708 bp in length  
\* 3143: contig of 712 bp in length  
\* 3244: gap of 100 bp  
\* 3245: contig of 723 bp in length  
\* 3967: contig of 723 bp in length  
\* 4066: gap of 100 bp  
\* 4068: contig of 724 bp in length  
\* 4792: gap of 100 bp  
\* 4892: contig of 708 bp in length  
\* 5599: contig of 708 bp in length  
\* 5600: gap of 100 bp  
\* 5700: contig of 726 bp in length  
\* 6425: contig of 726 bp in length  
\* 6426: gap of 100 bp  
\* 6526: contig of 694 bp in length  
\* 7219: gap of 100 bp  
\* 7319: contig of 727 bp in length  
\* 7320: contig of 718 bp in length  
\* 8038: gap of 100 bp  
\* 8138: contig of 709 bp in length  
\* 8846: contig of 709 bp in length  
\* 8847: gap of 100 bp  
\* 8947: contig of 718 bp in length  
\* 9665: gap of 100 bp  
\* 9765: contig of 727 bp in length  
\* 10491: contig of 718 bp in length  
\* 10592: gap of 100 bp  
\* 11316: contig of 725 bp in length  
\* 11317: gap of 100 bp  
\* 11416: contig of 710 bp in length  
\* 12126: gap of 100 bp  
\* 12226: contig of 717 bp in length  
\* 12227: contig of 718 bp in length  
\* 12943: contig of 717 bp in length  
\* 13043: gap of 100 bp  
\* 13044: contig of 712 bp in length  
\* 13755: contig of 745 bp in length  
\* 13756: gap of 100 bp  
\* 13855: contig of 704 bp in length  
\* 14556: gap of 100 bp  
\* 14556: contig of 745 bp in length  
\* 14660: contig of 745 bp in length

\* 15405: gap of 100 bp  
\* 15504: gap of 100 bp  
\* 15505: contig of 724 bp in length  
\* 15505: 16228: gap of 100 bp  
\* 16229: 16328: gap of 100 bp  
\* 16329: contig of 705 bp in length  
\* 17034: 17133: gap of 100 bp  
\* 17134: 17838: contig of 705 bp in length  
\* 17839: 17938: gap of 100 bp  
\* 17939: 18664: contig of 726 bp in length  
\* 18665: 18764: gap of 100 bp  
\* 18765: 19475: contig of 711 bp in length  
\* 19476: 19575: gap of 100 bp  
\* 19576: 20305: contig of 730 bp in length  
\* 20306: 20405: gap of 100 bp  
\* 20406: 21141: contig of 736 bp in length  
\* 21142: 21241: gap of 100 bp  
\* 21242: 21965: contig of 724 bp in length  
\* 21966: 22065: gap of 100 bp  
\* 22066: 22804: contig of 739 bp in length  
\* 22805: 22904: gap of 100 bp  
\* 22905: 23637: contig of 733 bp in length  
\* 23638: 23737: gap of 100 bp  
\* 23738: 24472: contig of 735 bp in length  
\* 24473: 24572: gap of 100 bp  
\* 24573: 25300: contig of 728 bp in length  
\* 25301: 25400: gap of 100 bp  
\* 25401: 26122: contig of 722 bp in length  
\* 26123: 26222: gap of 100 bp  
\* 26223: 26935: contig of 713 bp in length  
\* 26936: 27035: gap of 100 bp  
\* 27036: 27770: contig of 735 bp in length  
\* 27771: 27870: gap of 100 bp  
\* 27871: 28550: contig of 680 bp in length  
\* 28551: 28650: gap of 100 bp  
\* 28651: 29358: contig of 708 bp in length  
\* 29359: 29458: gap of 100 bp  
\* 29459: 30184: contig of 725 bp in length  
\* 30184: 30283: gap of 100 bp  
\* 30284: 31012: contig of 729 bp in length  
\* 31013: 31112: gap of 100 bp  
\* 31113: 31829: contig of 717 bp in length  
\* 31830: 31929: gap of 100 bp  
\* 31930: 32642: contig of 713 bp in length  
\* 32643: 32742: gap of 100 bp  
\* 32743: 33463: contig of 721 bp in length  
\* 33464: 33563: gap of 100 bp  
\* 33564: 34280: contig of 717 bp in length  
\* 34281: 34380: gap of 100 bp  
\* 34381: 35089: contig of 709 bp in length  
\* 35090: 35189: gap of 100 bp  
\* 35190: 35922: contig of 733 bp in length  
\* 35923: 36022: gap of 100 bp  
\* 36023: 36739: contig of 720 bp in length  
\* 36740: 36839: gap of 100 bp  
\* 36840: 37552: contig of 713 bp in length  
\* 37553: 37652: gap of 100 bp  
\* 37653: 38375: contig of 723 bp in length  
\* 38376: 38475: gap of 100 bp  
\* 38476: 39195: contig of 720 bp in length  
\* 39196: 39295: gap of 100 bp  
\* 39296: 40019: contig of 724 bp in length  
\* 40019: 41761: gap of 100 bp  
\* 41761: 42445: contig of 685 bp in length  
\* 42445: 42545: gap of 100 bp  
\* 42545: 43241: contig of 696 bp in length  
\* 43242: 43341: gap of 100 bp  
\* 43342: 44062: contig of 721 bp in length  
\* 44063: 44162: gap of 100 bp  
\* 44163: 44878: contig of 716 bp in length  
\* 44879: 44978: gap of 100 bp

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquery@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPPB; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-336A16 ls from the RCP1-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chorl.org/bacpac/home.htm>.

FEATURES

| source | Location/Qualifiers             |
|--------|---------------------------------|
|        | 1..71109                        |
|        | /organism="Mus musculus"        |
|        | /mol-type="genomic DNA"         |
|        | /db_xref="taxon:10090"          |
|        | /chromosome="2"                 |
|        | /clone="RP23-336A16"            |
|        | /clone.lib="RPCI-23"            |
|        | 21900 a 15578 c 14679 g 18952 t |
| ORIGIN |                                 |

Query Match 100.0% Score 231; DB 2; Length 70360;  
 Best Local Similarity 100.0%; Pred. No. 1..7e-65;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGAGGCTTAAGCAACTGCAAGGATGAGTCTGCTGAGCTTCCTGGAAAG 60  
 Db 31408 GTGGGTGAGGCTTAAGCAACTGCAAGGATGAGTCTGCTGAGCTTCCTGGAAAG 31349

QY 61 CACGCTGATTGATGAAGTTCAAGTGTCTAGTGTCCCTCCCCAGAAAAAGACAGC 120  
 Db 31348 CACGCTGATTGATGAAGTTCAAGTGTCTAGTGTCCCTCCCCAGAAAAAGACAGC 31289

QY 121 GGAGCTGGCCAAGAACCTCTCGGTGATGAGCTTAAGGATGCTTGTAGGGCCCT 180  
 Db 311288 GGAGCTGGCCAAGAACCTCTCGGTGATGAGCTTAAGGATGCTTGTAGGGCCCT 31229

QY 181 ATGTTGACACACTGGGTCAGACTTACCCATGAGCTTGCTT 231  
 Db 31228 ATGCTTGACACACTGGGTCAGACTTACCCATGAGCTTGCTT 31178

RESULT 10  
 AL935271/C Locus AL935271 71109 bp DNA linear ROD 05-MAR-2003  
 DEFINITION Mouse DNA sequence from clone RP23-336A16 on chromosome 2, complete sequence.  
 ACCESSION AL935271  
 VERSION HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE Collins P.  
 AUTHORS Direct Submission  
 TITLE Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 ACCESSION AL935271.9 GI:28874648  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 COMMENT Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

RESULT 11  
 BX294125/c Locus BX294125 77127 bp DNA linear ROD 17-APR-2003  
 DEFINITION Mouse DNA sequence from clone RP24-129F5 on chromosome 4, complete sequence.  
 ACCESSION BX294125  
 VERSION BX294125.5 GI:30024367  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus (house mouse)  
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Wallis,J.  
TITLE Direct Submission  
JOURNAL Submitted (17-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail inquiries:  
humquery@sanger.ac.uk Clone requests: clone@request@sanger.ac.uk  
COMMENT Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found there are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, ENSEMBL, SWISSPROT, Trl, TREMBL, Wp, WormPEP, Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C\_elegans/wormpep/RD24-129F5.1s from a Male (C57BL/6J) mouse BAC Library VECTOR: ptARBAC1.

FEATURES source  
1. 77127  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RD24-129F5"  
/clone\_id="RP01-24"  
from a Male (C57BL/6J) mouse BAC Library VECTOR: ptARBAC1.

BASE COUNT  
ORIGIN  
Query Match 21919 a 15230 c 15167 g 24811 t  
Best Local Similarity 100.0%; Score 231; DB 10; Length 77127;  
Matches 231; Pred. No. 1.7e-65; Conserv. 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGGCGAGGCCTAGACACTGAGAGGATAGCTTGCTGTGGAGG 60  
Db 32286 GGGGGCGAGGCCTAGACACTGAGAGGATAGCTTGCTGTGGAGG 32327

Qy 61 CACGTCATGATGCCATGAGGTTCAGTCCTGCTCCAGAAAAACGAAACG 120  
Db 32236 CACGTCATGATGCCATGAGGTTCAGTCCTGCTCCAGAAAAACGAAACG 322267

Qy 121 GGAGCTGGCCAAGAACCTCTCGGGTAGGCTAAGGGATGGTTGTGAGGCCCT 180  
Db 32286 GGAGCTGGCCAAGAACCTCTCGGTAGGCTAAGGGATGGTTGTGAGGCCCT 32156

Qy 181 ATGCTTGACACTGGGATAGACCTCTACCTTCACCATAGGGTGTGCTT 231  
Db 32206 ATGCTTGACACTGGGATAGACCTCTACCTTCACCATAGGGTGTGCTT 32156

RESULT 12  
AF481949/c AF481949 84250 bp DNA linear ROD 23-Oct-2002  
LOCUS Mus musculus Dom1 (Dom1) and Dom2 (Dom2) genes, exons 1 through 5  
DEFINITION Mus musculus Dom1 (Dom1) and Dom2 (Dom2) genes, exons 1 through 5  
/product="Dom2"

ACCESSION AF481949  
VERSION AF481949.1  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 84250)  
AUTHORS Berger,F.G.; Barbour,K.W.; Wei,F.-S.; Brannan,C.; Flotte,T.R.; Baumann,H. and Berger,F.G.  
TITLE The Murine alpha1-Proteinase Inhibitor Gene Family: Polymorphism, Chromosomal Location, and Structure  
JOURNAL Genomics 80 (5): 515-522 (2002)  
REFERENCE 2 (bases 1 to 84250)  
AUTHORS Tennant,M.; Brannan,C.; Wei,F.-S. and Flotte,T.R.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2002) Genetics Institute, University of Florida, 1600 SW Archer Rd, Gainesville, FL 32610-0266, USA  
REFERENCE 3 (bases 1 to 84250)  
AUTHORS Tennant,M.; Brannan,C.; Wei,F.-S. and Flotte,T.R.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUN-2002) Genetics Institute, University of Florida, 1600 SW Archer Rd, Gainesville, FL 32610-0266, USA  
REMARK Amino acid sequence updated by submitter  
FEATURES Location/Qualifiers  
1. 84250  
source  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="129/J"  
/db\_xref="taxon:10090"  
/chromosome="12"  
17789 .27751  
/gene="Dom1"  
join(17789..17817,23281..23912,25372..27474..276672,27474..27665)  
exon  
/gene="Dom1"  
23281..23912  
/gene="Dom1"  
/number=2  
join(23285..23912,25372..25642,26522..26672,27474..27665)  
CDS  
/gene="Dom1"  
/codon\_start=1  
/product="Dom1"  
/protein\_id="AA47468.1"  
/db\_xref="GI:21322147"  
/translation="WPSISWGILLAGLCLVPSFLAEDVQETDSQRDQSPLASHEI  
ATNLGDAISLYREIWHQSTNIPSPYATAMLSIGSKGDHTQLEGLQFL  
TOTSEADLRSFOHLLQTLNRPDSSELQLSNGNGLFVNNDLKLVEKFLPRAKHYOAEV  
FSVNPAASEPAKKVNDNEVKGTQCKIAEAVKQDQDTFLALNYLFGKWKDP  
ENTEAEFHDESTIVKVPOMTSLSDMDYHCSLTSWLLMDAGNAVAYFLPDG  
KMQHIEQTLKELIKFLLNRRRLQIAQHPRLSISGEYNLKTLMSPUGITLRFNNGA  
DLSGTTEENRPLKLSQAVKAVLTIDETGTEAAAATVQLQMPMSMPPIRLDFHPLFI  
IEEETQSPLIFVGKVVDPTHK"  
25372..25642  
/gene="Dom1"  
/number=3  
26522..26672  
/gene="Dom1"  
/gene="Dom2"  
/number=4  
27474..27751  
/gene="Dom1"  
/number=5  
65543..75506  
/gene="Dom2"  
/gene="Dom2"  
/product="Dom2"

join(65543..65571,71054..71685,73128..73398,74282..74432,  
75229..75506)

Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquery@sanger.ac.uk

---

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as gaps and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-400M24 1s from the RPCI-3 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

| FEATURES              | SOURCE       | LOCATIONS   | QUALIFIERS   |  |  |
|-----------------------|--------------|---|--|--|--|
|                       |              | 1..98653  | /organism="Mus musculus"<br>/mol_type="Genomic DNA"<br>/db_xref="taxon:10090"<br>/chromosome="4"<br>/clone_11D="RPCI-23"<br>/clone_11B="RPCI-23" |  |  |
|                       |              | 28064..a  | BASE COUNT   |  |  |
|                       |              | 22820..c  | ORIGIN   |  |  |
|                       |              | 25188..t  |  |  |  |
| Query Match           |              | 100..0%   | Query Match  | 100..0%  |  |
| Best Local Similarity |              | 100..0%   | Best Local Similarity  | 100..0%  |  |
| Matches               |              | 231;  | Matches  | 231;   |  |
| Conservative          |              | Mismatches 0;   | Conservative   | Mismatches 0;  |  |
|                       |              | Indels 0;   |  | Indels 0;  |  |
|                       |              | Gaps 0;   |  | Gaps 0;  |  |
| QY                    | 1            | GTGGGGTGGAGGCTAAGGACTGCACAGAGATAAGCTGCTGTTGGAGG 60        | QY   | 1  | GTGGGGTGGAGGCTAAGGACTGCACAGAGATAAGCTGCTGTTGGAGG 60     |
| Db                    | 8605         | GTGGGGTGGAGGCTAAGGACTGCACAGAGATAAGCTGCTGTTGGAGG 8546      | Db   | 43322  | GTGGGGTGGAGGCTAAGGACTGCACAGAGATAAGCTGCTGTTGGAGG 43381  |
| QY                    | 61           | CACTCTGATTGCAAGAGCTTCAGTGTCTAGTCCCTCCCAGAAAACGACAG 120    | QY   | 61   | CACGTCTGATTGCAAGACCTCTGCTTGTAGGGCCCT 120               |
| Db                    | 8545         | CACTCTGATTGCAAGAGCTTCAGTGTCTAGTCCCTCCCAGAAAACGACAG 8486   | Db   | 43382  | CACGTCTGATTGCAAGACCTCTGCTTGTAGGGCCCT 43441             |
| QY                    | 121          | GGAGCTGCCAAAGACCTCTCTGGTGTAGGGATGCTGTTGTAGGGCCCT 180      | QY   | 121  | GGAGCTGCCAAAGACCTCTCTGGTGTAGGGATGCTGTTGTAGGGCCCT 180   |
| Db                    | 8485         | GGAGCTGCCAAAGACCTCTCTGGTGTAGGGCCCT 8426                   | Db   | 43442  | GGAGCTGCCAAAGACCTCTCTGGTGTAGGGATGCTGTTGTAGGGCCCT 43501 |
| QY                    | 181          | ATGCTTGCAACTGGGATGAGCTCACCTCACCATGGCTTGTAGGGCCCT 231      | QY   | 181  | ATGCTTGCAACTGGGATGAGCTCACCTCACCATGGCTTGTAGGGCCCT 231   |
| Db                    | 8425         | ATGCTTGCAACACTGGGATGAGCTCACCTCACCATGGCTTGTAGGGCCCT 8375   | Db   | 43502  | ATGCTTGCAACTGGGATGAGCTCACCTCACCATGGCTTGTAGGGCCCT 43552 |
| RESULT 13             |              |   | RESULT 14  |  |  |
| AL731664              | AL731664     | 98653 bp linear ROD 24-JUN-2002                           | AC132401..0/c  | WCOMMENT   |  |
| LOCUS                 |              | sequence from clone RP23-400M24 on chromosome 4, complete | Sequence split into 5 fragments  | LOCUS AC132401 Accession AC132401  |  |
| DEFINITION            |              | sequence.   | Fragment Name  | Begin End  |  |
| ORGANISM              | Mus musculus | Mus musculus (house mouse)                                | AC132401..0  | 1 110000   |  |
| KEYWORDS              |              |   | AC132401..1  | 100001 210000  |  |
| REFERENCE             |              |   | AC132401..2  | 200001 310000  |  |
| AUTHORS               |              |   | AC132401..3  | 300001 410000  |  |
| TITLE                 |              |   | AC132401..4  | 400001 488146  |  |
| JOURNAL               |              |   |  | DNA 488146 bp  |  |
| COMMENT               |              |   |  | hg1 linear   |  |
|                       |              |   |  | HTG 03-SEP-2002  |  |
|                       |              |   |  | WORKING DRAFT  |  |
|                       |              |   |  | SEQUENCE, 140 unordered pieces.  |  |
|                       |              |   |  | AC132401   |  |
|                       |              |   |  | AC132401..1 to 98653   |  |
|                       |              |   |  | Lovell,J.  |  |
|                       |              |   |  | Submitted (22-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. E-mail: <a href="mailto:clonerequest@sanger.ac.uk">clonerequest@sanger.ac.uk</a> . Clone requests: <a href="mailto:clonerequest@sanger.ac.uk">clonerequest@sanger.ac.uk</a> . On Jun 26, 2002 this sequence version replaced gi:21540117. |  |
|                       |              |   |  | Genome Center  |  |

VERSION AC132401.1 GI:22657894  
 KEYWORDS HTG; HTGS; PHASEI; HRGS; DRAFT.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 49916)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 JOURNAL The sequence of *Mus musculus* clone  
 JOURNAL Unpublished  
 JOURNAL 2 (bases 1 to 489146)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 JOURNAL Direct Submission (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 COMMENT -----  
 ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
 ----- Project Information -----  
 Center project name: M\_Ba040A08  
 ----- Summary Statistics -----  
 Sequencing vector: M13; Ø%  
 Sequencing vector: Plasmid; 100%  
 Chemistry: Dye primer EM; Ø% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus Quality: 430636 bases at least Q40  
 Consensus Quality: 448443 bases at least Q30  
 Consensus Quality: 460211 bases at least Q20  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 140 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1212: contig of 1212 bp in length  
 \* 1312: gap of unknown length  
 \* 1313: contig of 1265 bp in length  
 \* 2578: 2677: gap of unknown length  
 \* 2678: 4409: gap of 1732 bp in length  
 \* 4410: 4509: gap of unknown length  
 \* 4510: 5692: contig of 1183 bp in length  
 \* 5693: 5792: gap of unknown length  
 \* 5793: 6999: contig of 1207 bp in length  
 \* 7000: 7099: gap of unknown length  
 \* 7100: 8301: contig of 1202 bp in length  
 \* 8302: 8401: gap of unknown length  
 \* 8402: 9700: contig of 1299 bp in length  
 \* 9701: 9800: gap of unknown length  
 \* 9801: 10916: contig of 1116 bp in length  
 \* 10917: 11016: gap of unknown length  
 \* 11017: 12583: contig of 1567 bp in length  
 \* 12584: 12683: gap of unknown length  
 \* 12684: 13755: contig of 1072 bp in length  
 \* 13756: 13855: gap of unknown length  
 \* 13856: 15575: contig of 1720 bp in length  
 \* 15576: 15675: gap of unknown length  
 \* 15676: 16894: contig of 1219 bp in length  
 \* 16895: 16994: gap of unknown length  
 \* 16995: 18323: contig of 1329 bp in length  
 \* 18324: 18423: gap of unknown length  
 \* 18424: 19659: contig of 1236 bp in length  
 \* 19659: 19759: gap of unknown length  
 \* 19760: 21346: contig of 1587 bp in length  
 \* 21347: 21446: gap of unknown length  
 \* 22722: contig of 1276 bp in length  
 -----  
 22723 22822: gap of unknown length  
 22823 23882: contig of 1060 bp in length  
 23883 23982: gap of unknown length  
 23983 25215: contig of 1233 bp in length  
 25315: gap of unknown length  
 25316 26372: contig of 1057 bp in length  
 26472: gap of unknown length  
 26473 27911: contig of 1439 bp in length  
 27912 28011: gap of unknown length  
 28012 29838: contig of 1827 bp in length  
 29839 29938: gap of unknown length  
 29939 31555: contig of 1617 bp in length  
 31556 31655: gap of unknown length  
 31656 33192: contig of 1537 bp in length  
 33193 33292: gap of unknown length  
 33293 34857: contig of 1565 bp in length  
 34858 34957: gap of unknown length  
 34958 36321: contig of 1364 bp in length  
 36322 36421: gap of unknown length  
 36422 38182: contig of 1760 bp in length  
 38182 38281: gap of unknown length  
 38282 39578: contig of 1297 bp in length  
 39579 39678: gap of unknown length  
 39679 41309: contig of 1631 bp in length  
 41310 41409: gap of unknown length  
 41410 42539: contig of 1130 bp in length  
 42540 42639: gap of unknown length  
 42640 44944: contig of 2305 bp in length  
 44945 45045: gap of unknown length  
 45045 46293: contig of 1249 bp in length  
 46294 46393: gap of unknown length  
 46394 47882: contig of 1489 bp in length  
 47883 47982: gap of unknown length  
 47983 49351: contig of 1362 bp in length  
 49352 49451: gap of unknown length  
 49452 50966: contig of 1515 bp in length  
 50967 51066: gap of unknown length  
 51067 52365: contig of 1299 bp in length  
 52366 52465: gap of unknown length  
 52466 54067: contig of 1602 bp in length  
 54068 54167: gap of unknown length  
 54168 55875: contig of 1707 bp in length  
 55875 55974: gap of unknown length  
 55975 57285: contig of 1311 bp in length  
 57286 59623: gap of 2238 bp in length  
 59624 59724: gap of unknown length  
 59724 61001: contig of 1278 bp in length  
 61002 61101: gap of unknown length  
 61102 62734: contig of 1633 bp in length  
 62735 62834: gap of unknown length  
 62835 64265: contig of 1431 bp in length  
 64266 64365: gap of unknown length  
 64366 66070: contig of 1705 bp in length  
 66071 66171: gap of unknown length  
 66171 68449: contig of 2279 bp in length  
 68450 68549: gap of unknown length  
 68550 69974: contig of 1315 bp in length  
 69975 70074: gap of unknown length  
 70075 71868: contig of 1794 bp in length  
 71869 71968: gap of unknown length  
 71969 73384: contig of 1416 bp in length  
 73385 73484: gap of unknown length  
 73485 74799: contig of 1425 bp in length  
 74799 74899: gap of unknown length  
 74800 74900: contig of 2499 bp in length  
 74900 77398: contig of 2499 bp in length  
 77399 77498: gap of unknown length  
 77499 78516: contig of 1018 bp in length  
 78517 78616: gap of unknown length  
 78616 80110: contig of 1494 bp in length  
 80111 80210: gap of unknown length  
 80211 81382: contig of 1172 bp in length  
 81383 81482: gap of unknown length

Query Match 100.0% Score 231; DB 2; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 1.e-65;  
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAAGCACTGCTAGTGGCTAGTCCTGTGGAGG 60  
 Db 35877 GTGGGTGCGAGGCTAAAGCACTGCTAGTGGCTAGTCCTGTGGAGG 35818

QY 61 CACCTCTGATTGATGAAGCTCTGGCTAAGGGATGGTTTGTAAGGCCCT 120  
 Db 35817 CACCTCTGATTGATGAAGCTCTGGCTAAGGGATGGTTTGTAAGGCCCT 35758

QY 121 GGAGCTGGCAAGACCTCTGGCTAAGGGATGGTTTGTAAGGCCCT 180  
 Db 35757 GGAGCTGGCAAGACCTCTGGCTAAGGGATGGTTTGTAAGGCCCT 35698

QY 181 ATGCTTGACACCTGGATCAGACCTCACCCATGAGCTTGCT 231  
 Db 35697 ATGCTTGACACCTGGATCAGACCTCACCCATGAGCTTGCT 35647

RESULT 15  
 AL29223\_2/c  
 WPCOMMENT Sequence split into 6 fragments Locus AL929223 Accession AL929223

| Fragment Name | Begin  | End    |
|---------------|--------|--------|
| AL29223_0     | 1      | 100001 |
| AL29223_1     | 100001 | 210000 |
| AL29223_2     | 200001 | 310000 |
| AL29223_3     | 300001 | 410000 |